Pharmacoepidemiology

Pharmacoepidemiology

EDITED BY

Brian L. Strom MD, MPH

George S. Pepper Professor of Public Health and Preventive Medicine
Professor of Biostatistics and Epidemiology, of Medicine, and of Pharmacology
Chair, Department of Biostatistics and Epidemiology
Director, Center for Clinical Epidemiology and Biostatistics
Vice Dean for Institutional Affairs
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA, USA

Stephen E. Kimmel MD, MSCE

Professor of Medicine and of Epidemiology
Senior Scholar, Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA, USA

Sean Hennessy PHARMD, PHD

Associate Professor of Epidemiology and of Pharmacology
Senior Scholar, Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA, USA

FIFTH EDITION

WILEY-BLACKWELL

A John Wiley & Sons, Ltd., Publication

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office: John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell

Library of Congress Cataloging-in-Publication Data
Pharmacoepidemiology / edited by Brian L. Strom, Stephen E. Kimmel, Sean Hennessy. – 5th ed.
p. ; cm.
Includes bibliographical references and index.
ISBN-13: 978-0-470-65475-0 (hard cover : alk. paper)
ISBN-10: 0-470-65475-9 (hard cover : alk. paper)
1. Pharmacoepidemiology. 2. Pharmacology. I. Strom, Brian L. II. Kimmel, Stephen E. III. Hennessy, Sean.
[DNLM: 1. Pharmacoepidemiology–methods. QZ 42]
RM302.5.P53 2012
615'.7042–dc23
2011019285

A catalogue record for this book is available from the British Library.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Set in 9/12pt Meridien by Toppan Best-set Premedia Limited, Hong Kong

1 2012

Contents

List of Contributors

Trisha Acri
Assistant Professor
Department of Family and Community Medicine
Temple University School of Medicine
Philadelphia, PA
USA

Susan E. Andrade
Senior Research Associate and Research Associate Professor
Meyers Primary Care Institute
and
University of Massachusetts Medical School
Worcester, MA
USA

Elizabeth Andrews
Vice President
Pharmacoepidemiology and Risk Management
Research Triangle Institute Health Solutions
Research Triangle Park, NC
USA

Peter Arlett
Head
Pharmacovigilance and Risk Management
European Medicines Agency
London, UK

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

Jeffrey S. Barrett
Director
Laboratory for Applied Pharmacokinetics and Pharmacodynamics
Director
Pediatric Pharmacology Research Unit
The Children's Hospital of Philadelphia
Research Professor, Pediatrics
Kinetic Modeling and Simulation (KMAS) Core Director
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

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

Ulf Bergman
Professor, Senior Medical Officer
Division of Clinical Pharmacology
Department of Laboratory Medicine
Karolinska Institute
WHO Collaborating Centre for Drug Utilization Research and Clinical Pharmacological Services
and
Centre for Pharmacoepidemiology
Karolinska University Hospital-Huddinge
Stockholm, Sweden

Jesse A. Berlin
Vice President
Department of Epidemiology
Johnson & Johnson Pharmaceutical Research and Development
Titusville, NJ
USA

Stella Blackburn
EMA Risk Management Development and Scientific Lead
European Medicines Agency
London, UK

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

Denise Boudreau
Scientific Investigator
Group Health Research Institute
Seattle, WA
USA

Arthur Caplan
Director
Center for Bioethics
Professor of Medical Ethics
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

M. Soledad Cepeda
Director
Department of Epidemiology
Johnson & Johnson Pharmaceutical Research and Development
Titusville, NJ
USA

Robert T. Chen
HIV Vaccine and Special Studies Team Leader
Division of HIV/AIDS Prevention
Centers for Disease Control and Prevention
Atlanta, GA
USA

Patricia F. Coogan
Associate Professor of Epidemiology
Slone Epidemiology Center
Boston University
Boston, MA
USA

Francesca Cunningham
Director
Center for Medication Safety
and
Program Manager
Outcomes Research PBM Services
Department of Veterans Affairs
Center for Medication Safety
Hines, IL
USA

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

Gregory W. Daniel
Vice President
Government and Academic Research
Healthcore
Alexandria, VA
USA

Hassy Dattani
Research Director
Cegedim Strategic Data Medical Research Ltd
London, UK

Robert L. Davis
Director of Research
Center for Health Research—Southeast
Kaiser Permanente Georgia
Atlanta, GA
USA

Nancy A. Dreyer
Chief of Scientific Affairs
Outcome Sciences Inc.
Cambridge, MA
USA

Antoine C. El Khoury
Leader, Outcomes Research
Global Health Outcomes
Merck & co., Inc
West Point, PA
USA

Chris Feudtner
Associate Professor
Department of Pediatrics
The Children's Hospital of Philadelphia
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Brian T. Fisher
Assistant Professor of Pediatrics
Center for Pediatric Clinical Effectiveness
The Children's Hospital of Philadelphia
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Cristin Palumbo Freeman
Research Project Manager
Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Nicolle M. Gatto
Senior Director
Epidemiology
Worldwide Safety Strategy
Pfizer Inc.
New York, NY
USA

Joel M. Gelfand
Medical Director, Clinical Studies Unit
Assistant Professor of Dermatology and Epidemiology
Senior Scholar, Center for Clinical Epidemiology and Biostatistics
Departments of Dermatology and Biostatistics and Epidemiology
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

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

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

Henry A. Glick
Professor of Medicine
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Robert Gross
Associate Professor of Medicine and Epidemiology
Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

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

Gordon H. Guyatt
Professor
Department of Clinical Epidemiology and Biostatistics
McMaster University
Health Sciences Center
and
Department of Medicine
St Joseph's Hospital
Hamilton
Ontario, Canada

Katherine Haffenreffer
Project Administrator
Harvard Pilgrim Health Care Institute
and
Department of Population Medicine
Harvard Medical School
Boston, MA
USA

Sean Hennessy
Associate Professor of Epidemiology and of Pharmacology
Senior Scholar, Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Ron M.C. Herings
Director
PHARMO Institute
Utrecht
The Netherlands
and
Associate Professor of Pharmacoepidemiology
Department of Health Policy and Management
Erasmus University Rotterdam
Rotterdam, The Netherlands

Roman Jaeschke
Professor
Department of Clinical Epidemiology and Biostatistics
McMaster University
Health Sciences Center
and
Professor
Department of Medicine
St Joseph's Hospital
Hamilton
Ontario, Canada

Bradley C. Johnston
Assistant Professor
Department of Clinical Epidemiology and Biostatistics
McMaster University
Health Sciences Center
Hamilton
Ontario, Canada

Judith K. Jones
President and CEO
The Degge Group Ltd
Arlington, VA
USA
and
Adjunct Professor and Lecturer
University of Michigan School of Public Health Summer Program
Georgetown University
Washington, DC
USA

Jason Karlawish
Professor of Medicine
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Claudia Karwoski
Director
Division of Risk Management
Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MA
USA

David W. Kaufman
Associate Director
Slone Epidemiology Center at Boston University
and
Professor of Epidemiology
Boston University School of Public Health
Boston, MA
USA

Aaron S. Kesselheim
Assistant Professor of Medicine
Division of Pharmacoepidemiology and Pharmacoeconomics
Department of Medicine
Brigham and Women's Hospital
Harvard Medical School
Boston, MA
USA

Carin J. Kim
Mathematical Statistician
Center for Drug Evaluation and Research
US Food and Drug Administration
Silver Spring, MD
USA

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

Karel Kostev
Senior Research Analyst
Centre of Excellence Patient Data
IMS Health GmbH & Co OHG
Frankfurt/Main, Germany

Martin Kulldorff
Professor, Biostatistician
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute
Boston, MA
USA

Sinéad M. Langan
National Institute for Health Research Clinician Scientist and Honorary Consultant Dermatologist
London School of Hygiene and Tropical Medicine
and St John's Institute of Dermatology
London, UK

Ebbing Lautenbach
Professor of Medicine and Epidemiology
Senior Scholar, Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Deborah Layton
Principal Research Fellow
Honorary Senior Lecturer
Portsmouth University
Drug Safety Research Unit academic contact for collaborative programme MSc in Pharmacovigilance
Portsmouth University
Portsmouth, UK

David Lee
Director
Technical Strategy and Quality
Center for Pharmaceutical Management
Management Sciences for Health, Inc.
Arlington, VA
USA

Jacques LeLorier
Professor
Department of Medicine
Department of Pharmacology
Université de Montréal
CHUM/Centre de Researche Hôpital Hôtel-Dieu de Montréal, Montreal, Quebec, Canada

Samuel M. Lesko
Medical Director and Director of Research
Northeast Regional Cancer Institute
Scranton, PA
USA
and
Adjunct Professor of Public Health Sciences
Pennsylvania State University College of Medicine
Hershey, PA
USA
and
Adjunct Professor of Basic Sciences
The Commonwealth Medical College
Scranton, PA
USA

Hubert G. Leufkens
Professor
Department of Pharmacoepidemiology and Clinical Pharmacotherapy
Utrecht Institute for Pharmaceutical Sciences
Utrecht University
Utrecht, The Netherlands

Mitchell Levine
Professor
Department of Clinical Epidemiology & Biostatistics, Department of Medicine
McMaster University
Centre for Evaluation of Medicines
Hamilton
Ontario, Canada

Peter K. Lindenauer
Director
Center for Quality of Care Research
Baystate Medical Center
Springfield
and Associate Professor
Department of Medicine
Tufts University School of Medicine
Boston, MA
USA

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

Darren R. Linkin
Assistant Professor of Medicine
Associate Scholar, Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine at the University of Pennsylvania
and
Philadelphia VA Medical Center
Philadelphia, PA
USA

Helene Levens Lipton
Professor of Pharmacy and Health Policy
Schools of Medicine and Pharmacy
University of California at San Francisco
San Francisco, CA
USA

Sumit R. Majumdar
Professor of Medicine
Department of Medicine
Faculty of Medicine and Dentistry
University of Alberta
Edmonton
Alberta, Canada

Danica Marinac-Dabic
Director
Division of Epidemiology
Office of Surveillance and Biometrics
Center for Devices and Radiological Health
US Food and Drug Administration
Silver Spring, MD
USA

the late Kenneth L. Melmon
Stanford University School of Medicine
Stanford, CA
USA

Colleen J. Metge
Associate Professor (Senior Scholar)
Faculty of Pharmacy
University of Manitoba
Winnipeg
Manitoba, Canada

Allen A. Mitchell
Director
Slone Epidemiology Center at Boston University
and
Professor of Epidemiology and Pediatrics
Boston University Schools of Public Health and Medicine
Boston, MA
USA

Jingping Mo
Senior Director
Epidemiology, Worldwide Safety Strategy
Pfizer Inc.
New York, NY
USA

Yola Moride
Associate Professor
Faculty of Pharmacy
Université de Montréal
and
Researcher
Research Centre
University of Montreal Hospital Centre (CRCHUM)
Montreal, Canada

Sharon-Lise Normand
Professor of Health Care Policy (Biostatistics)
Department of Health Care Policy
Harvard Medical School
and
Professor
Department of Biostatistics
Harvard School of Public Health
Boston, MA
USA

Alexis Ogdie
Instructor in Medicine
Division of Rheumatology
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Julie R. Palmer
Professor of Epidemiology
Slone Epidemiology Center
Boston University
Boston, MA
USA

John Parkinson
Director
General Practice Research Database (GPRD)
Medicines and Healthcare products Regulatory Agency (MHRA)
London, UK

Pamala A. Pawloski
Research Investigator
HealthPartners Research Foundation
Bloomington, MN
USA

Lars Pedersen
Associate Professor of Clinical Epidemiology
Department of Clinical Epidemiology
Aarhus University Hospital
Aarhus, Denmark

Richard Platt
Professor and Chair
Department of Population Medicine,
Harvard Medical School
and
Executive Director
Harvard Pilgrim Health Care Institute
Boston, MA
USA

Daniel Polsky
Professor of Medicine
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Charles Poole
Associate Professor
Department of Epidemiology
Gillings School of Global Public Health
University of North Carolina
Chapel Hill, NC
USA

Judith A. Racoosin
Sentinel Initiative Scientific Lead
Office of Medical Policy
Center for Drug Evaluation and Research
Food and Drug Administration
Silver Spring, MD
USA

Marsha A. Raebel
Investigator
Institute for Health Research
Kaiser Permanente Colorado
and
Clinical Professor
School of Pharmacy
University of Colorado at Denver
Denver, CO
USA

Timothy R. Rebbeck
Professor of Epidemiology
Senior Scholar, Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Shelby D. Reed
Associate Professor in Medicine
Duke University
Durham, NC
USA

Robert F. Reynolds
Vice President
Epidemiology, Worldwide Safety Strategy
Pfizer Inc.
New York, NY
USA

Mary Elizabeth Ritchey
Associate Division Director
Food and Drug Administration
Center for Devices and Radiological Health
Silver Spring, MD
USA

Melissa A. Robb
Sentinel Initiative Program Director
Office of Medical Policy
Center for Drug Evaluation and Research
Food and Drug Administration
Silver Spring, MD
USA

Lynn Rosenberg
Professor of Epidemiology
Associate Director
Slone Epidemiology Center
Boston University
Boston, MA
USA

Rita Schinnar
Senior Research Project Manager and Analyst
Center for Clinical Epidemiology and Biostatistics
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Sebastian Schneeweiss
Associate Professor of Medicine and Epidemiology
Harvard Medical School
and
Vice Chief
Division of Pharmacoepidemiology
Department of Medicine
Brigham & Women's Hospital
Boston, MA
USA

Kevin A. Schulman
Professor of Medicine and Gregory Mario and Jeremy Mario Professor of Business Administration
Duke University
Durham, NC
USA

Holger Schünemann
Professor and Chair
Department of Clinical Epidemiology and Biostatistics
McMaster University
Health Sciences Center
and
Professor
Department of Medicine
St Joseph's Hospital
Hamilton
Ontario, Canada

Art Sedrakyan
Associate Professor
Director
Comparative Effectiveness Program at HSS and NYP
Weill Cornell Medical College
New York, NY
USA

John Seeger
Lecturer in Medicine
Brigham and Women's Hospital/ Harvard Medical School
Boston, MA
USA

Hanna M. Seidling
Head of Cooperation Unit Clinical Pharmacy
Department of Clinical Pharmacology and Pharmacoepidemiology
Cooperation Unit Clinical Pharmacy
University of Heidelberg
Heidelberg
Germany

Saad A.W. Shakir
Professor and Director
Drug Safety Research Unit
Southampton, UK

Rachel E. Sherman
Director, Office of Medical Policy
Center for Drug Evaluation and Research
Food and Drug Administration
Silver Spring, MD
USA

Betsy L. Sleath
Professor and Chair
Division of Pharmaceutical Outcomes and Policy
Eshelman School of Pharmacy
University of North Carolina at Chapel Hill
Chapel Hill, NC
USA

Rachel E. Sobel
Senior Director
Epidemiology
Worldwide Safety Strategy
Pfizer Inc.
New York, NY
USA

Stephen B. Soumerai
Professor of Population Medicine
Professor of Medicine and Director of the Drug Policy Research Group
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute
Boston, MA
USA

Brian L. Strom
George S. Pepper Professor of Public Health and Preventive Medicine
Professor of Biostatistics and Epidemiology, of Medicine, and of Pharmacology
Chair
Department of Biostatistics and Epidemiology
Director
Center for Clinical Epidemiology and Biostatistics
Vice Dean for Institutional Affairs
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
USA

Samy Suissa
James McGill Professor of Epidemiology, Biostatistics and Medicine
McGill University
and
Director
Centre for Clinical Epidemiology
Lady Davis Research Institute
Jewish General Hospital
Montreal
Quebec, Canada

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

Priscilla Velentgas
Director of Epidemiology
Outcome Sciences Inc.
Cambridge, MA
USA

Claudia Vellozzi
Deputy Director
Immunization Safety Office
Division of Healthcare Quality and Promotion
National Center for Emerging and Zoonotic Diseases
Centers for Disease Control and Prevention
Atlanta, GA
USA

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

Janet Woodcock
Director
Center for Drug Evaluation and Research
Food and Drug Administration
Silver Spring, MD
USA

Athena F. Zuppa
Associate Professor of Pediatrics, Anesthesia and Critical Care Medicine
Laboratory for Applied Pharmacokinetics and Pharmacodynamics
Children's Hospital of Philadelphia
Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA
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,
> *Medical Essays, 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, the 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 the combination of amoxicillin and 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. New in this fifth edition are progressive multifocal leukoencephalopathy from natalizumab; hepatotoxicity from pemoline and from lumiracoxib; serious cardiovascular complications from rosiglitazone, tegaserod, sibutramine, rimonabant, valdecoxib, pergolide, and propoxyphene; fatal adverse reactions when used with alcohol

from hydromorphone; and serious and sometimes fatal brain infections from efalizumab. Many 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.

As this edition goes to press, 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) is being doubled in size, FDA has been given new regulatory authority after drug marketing, and has also been 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 health-care system in the world. These developments portend major changes for our field.

The joining of the fields of clinical pharmacology and epidemiology resulted in the development of a new field: pharmacoepidemiology, the study of the use of and the 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 (ISPE), an early idea when the first edition of this book was written, has grown into a major international scientific force, with over 1280 members from 52 countries, an extremely successful annual meeting attracting close to 1000 attendees, a large number of very active committees and scientific interest groups, and its own journal (*Pharmacoepidemiology and Drug Safety*). 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 as well. We have also become institutionalized as a subfield within the field of clinical pharmacology, with a Pharmacoepidemiology Section within the American Society for Clinical Pharmacology and Therapeutics, recently reorganized into a Section on Drug Safety, 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, to have data in hand and available when questions arise, rather than waiting to begin to collect 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, and manufacturers are scrambling to respond. Requirements that a drug be proven to be cost-effective have been added to national, local, and insurance health-care 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 more pharmacoepidemiology manpower. Pharmacoepidemiologic research funding is now more plentiful, and even limited support for training is now available.

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. Now, years later however, the utility of drug utilization review programs is being questioned. In addition, the Joint Commission on Accreditation of Health Care Organizations now requires that every hospital in the country have 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 increased concern by the public about privacy has made pharmacoepidemiologic research much more difficult.

In the first edition, the goal was to help introduce this new field to the scientific world. The explosion in interest in the field, the rapid scientific progress that has been made, and the unexpected sales of the first edition led to the second edition. The continued maturation of what used to be a new 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 prepared a textbook version, which has been widely used. Now, six years after the fourth edition, the field continues to rapidly change, so it is time for a new edition. For this edition as well, we now include two co-editors who have both shared the work and contributed many important new ideas.

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

As in earlier editions, Part I of this book provides background information on what is included in the field of pharmacoepidemiology, a description of the study designs it uses, a description 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, regulatory agencies, and the law. 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 database. 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 the field. Our expectation is that Parts I, II, III, and VI of this book will be of greatest interest to those new to the field. In contrast, Parts III, IV, V, and VI should be of greatest interest to those with some background, who want a more in-depth view of the field.

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* may 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 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 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 field. 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 the field has seen. In the process, we hope is that it can continue to serve to assist the field in its development.

Brian L. Strom
Stephen E. Kimmel
Sean Hennessy
Philadelphia

Acknowledgements

There are many individuals and institutions to whom we owe thanks for their contributions to our efforts in preparing this book. Over the years, our pharmacoepidemiology work has been supported mostly by grants, contracts, and cooperative agreements from the US government, especially multiple different branches of the National Institutes of Health, the Agency for Healthcare Research and Quality, Food and Drug Administration, and the Department of Veterans Affairs. Other funders of our work include the American Cancer Society, the American College of Cardiology, the American College of Clinical Pharmacy Foundation, the Asia Foundation, the Charles A. Dana Foundation, the Joint Commission on Prescription Drug Use, the Pennsylvania Department of Health, the Rockefeller Foundation, the Andrew W. Mellon Foundation, and the International Clinical Epidemiology Network, Inc. We have also benefitted from project grants from Aetna, Alza Corporation, Amgen, AstraZeneca, Bayer Corporation, Bayer Consumer Care, Berlex Laboratories, Boran Pharmaceuticals, Bristol-Myers Squibb, the Burroughs Wellcome Company, Ciba-Geigy Corporation, COR Therapeutics Inc., GlaxoSmith-Kline, Glaxo-SmithKline Beecham, Glaxo Wellcome, Health Information Designs, Inc., Hoechst -Roussel Pharmaceuticals, Hoffman-La Roche, Inc., Integrated Therapeutics, Inc., a subsidiary of Schering-Plough Corporation, International Formula Council, Key Pharmaceuticals Inc., Marion Merrell Dow, Inc., McNeil Consumer Products, McNeil Pharmaceuticals, Mead Johnson Pharmaceuticals, Merck and Company, Novartis Pharmaceuticals Corp., Pfizer Inc, Pharming, PharMark Inc., A.H. Robins Company, Rowell Laboratories, Sandoz Pharmaceuticals, Schering Corporation, Searle Pharmaceutical, Shire, Smith Kline and French Laboratories, Sterling Winthrop Inc., Syntex, Inc., Takeda Pharmaceuticals North America, the Upjohn Company, and Wyeth-Ayerst Research. In addition, generous support to our pharmacoepidemiology training program has been provided by Abbott Laboratories, Alza Corporation, Amgen, Aventis Pharmaceuticals, Inc., Berlex Laboratories, Inc., Ciba-Geigy Corporation, Genentech, Inc., Hoechst-Marion-Roussel, Inc., Hoffman LaRoche, Integrated Therapeutics Group, Inc., Johnson and Johnson, Mary E Groff Charitable Trust, Merck and Company, Inc., McNeil Consumer Product Company, McNeil Consumer Healthcare, Novartis Pharmaceuticals Corporation, , Pfizer Inc., Sanofi Aventis, Sanofi Pasteur, SmithKline Beecham Pharmaceuticals, Whitehall-Robins Healthcare, and Wyeth-Ayerst Research. Finally, we would like to thank the University of Pennsylvania. While none of this support was specifically intended to support the development of this book, without this assistance, we would not have been able to support our careers in pharmacoepidemiology. Finally, we would like to thank our publisher, John Wiley & Sons, Ltd., for their assistance and insights, both in support of this book, and in support of the field's journal, *Pharmacoepidemiology and Drug Safety*.

Rita Schinnar's contributions to this book were immeasurable, encompassing both the role of Managing Editor and reviewing all of the chapters, editing them thoughtfully and posing substantive questions and issues for the authors to address. She also co-authored one chapter and assisted BLS with researching topics to update his other chapters. Catherine Vallejo assisted with early arrangements to contact the authors. Finally, Anne Saint John provided superb help in preparing both the manu-

scripts for my chapters and all of the other chapters for submission to Wiley.

BLS would like to thank Steve Kimmel and Sean Hennessy for joining him as co-editors in this edition. These are two very special and talented men. It has been BLS's pleasure to help to train them, now too many years ago, help them cultivate their own careers, and 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 parents for the support and education that were critical to his being able to be successful in his career. BLS would also like to thank Paul D. Stolley, M.D., M.P.H. and the late Kenneth L. Melmon, M.D., 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 20 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, patience, and willingness to engage in "no talking time" in the study while Dad worked.

SH also thanks BLS, his long-time 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

CHAPTER 1
What is Pharmacoepidemiology?

Brian L. Strom
Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, 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 what it had before. Although this has given health-care providers the ability to provide better medical care for their patients, it has also resulted 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 (ADRs), and 1.5 million US hospitalizations each year result from ADRs; yet, 20–70% of ADRs 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 many issues other than adverse reactions, as well.

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 reviewed, 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 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, Fifth Edition. Edited by Brian L. Strom, Stephen E. Kimmel, Sean Hennessy.

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 specific patient at hand. This individualization of therapy requires the determination of a risk/benefit ratio specific to the patient at hand. 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 his 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 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 that are the result of an exaggerated but otherwise usual pharmacologic effect of the drug, sometimes called *Type A reactions*, versus those that 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 the drug unusually slowly, leading to drug levels that are too high (see also Chapter 34). Third, they may have normal drug levels, but for some reason are overly sensitive to them (see Chapter 34).

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; see Chapter 34) 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 at other times this term is used to refer to all of pharmacoepidemiology). However, determining causation in case reports of adverse reactions can be problematic (see Chapter 33), as can attempts to compare the effects of drugs in the same class (see Chapter 32). This has led academic investigators, industry, FDA, and the legal com-

munity to turn to the field of epidemiology. Specifically, *studies of adverse effects* have been supplemented with *studies of adverse events* (ADEs). 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 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 performing 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 when it is first discovered or synthesized through 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 assure 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 sulfanilamide 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 the 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 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 United Kingdom, this resulted in the establishment in 1968 of the Committee on Safety of Medicines. Later, the World Health Organization 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 man. 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 below. 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 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, the so-called drug lag.[13] However, the drugs that are marketed 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 neurological 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, 5 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 non-steroidal 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 28). 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 with 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 are 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), grepafloxacin (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 26). Further, between 1990 and 2004, at least 15 non-cardiac drugs, including astemizole, cisapride, droperidol, grepafloxacin, halofantrine, pimozide, propoxyphene, rofecoxib, sertindole, sibutramine terfenadine, terodiline, thioridazine, levacetylmethadol, and ziprasidone, were subject to significant regulatory actions because of cardiac concerns.[96]

Since 1993, in 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), and as part of that system issuing monthly notifications of label changes. Compared to the 20 to 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]

According to a study 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 above, 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 led to 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, II, III, and 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, III, and 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 only $2.4 million in extramural safety research. This state of affairs has coincided with the growing numbers of drug crises cited above. 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.

However, there now was growing concern, in Congress and 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 FDA's regulatory mission, to complement the latter. For example, such a board could investigate drug safety crises such as those cited above, 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 FDA's Office of New Drugs and Office of Drug Safety to communicate with each other on safety issues, failure of FDA to track ongoing postmarketing studies, ambiguous role of FDA's Office of Drug Safety in scientific advisory committees, limited authority by FDA to require the pharmaceutical industry to perform studies to obtain needed data, concerns about culture problems at FDA where recommendations by members of the FDA's drug safety staff were not followed, and concerns about conflict of interest involving advisory committee members. This Institute of Medicine report was influential in shaping PDUFA IV.

Indeed, with the passage of PDUFA IV, FDA 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 29). 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 systems, including the MedWatch Plus Portal, the FDA Adverse Event Reporting System, the Sentinel System (a virtual national medical product safety system, see Chapter 30), 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]

Intellectual development of pharmacoepidemiology emerging from drug crises

Several developments of 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 the Kefauver–Harris Amendments required formal efficacy studies for old drugs that were approved earlier. These requirements created demand for new expertise and new methods. In addition, the mid-1960s saw the publication of a series of drug utilization studies.[112–116] These studies 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 24 and 25).

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[117] (see Chapter 19). 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.[118] The Computerized Online Medicaid Analysis and Surveillance System (COMPASS®) was first developed in 1977, using Medicaid billing data to perform pharmacoepidemiologic studies[119] (see Chapter 14). The Drug Surveillance Research Unit, now called the Drug Safety Research Trust, was developed in the United Kingdom in 1980, with its innovative system of Prescription–Event Monitoring[120] (see Chapter 20). Each of these represented major contributions to the field of pharmacoepidemiology. These and newer approaches are reviewed in Part III of this book.

In the examples of drug crises mentioned in the earlier section, these were serious but uncommon drug effects, and these experiences have led to an accelerated search for new methods to study drug effects in large numbers of patients. This led to a shift from adverse effect studies to adverse event studies, with 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,[121] 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,[122] which were updated in 2007.[123] Since the late 1990s, pharmacoepidemiologic research has also been increasingly burdened by concerns about patient confidentiality[124–128] (see also Chapter 35).

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 described earlier). Starting in 1999 and incrementally adding more centers in 2002, 2006, and 2007, the Agency for Healthcare Research and Quality (AHRQ) which was selected to administer this program has been funding up to 14 Centers for Education and Research and Therapeutics (CERTs)[129] (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. These CERTs centers 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 health-care items and services.[130] In response, AHRQ created in 2005 the DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) Network to support in academic settings the conduct of studies on effectiveness, safety, and usefulness of drugs and other treatments and services.[131]

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, starting in the late 1990s, there were new initiatives begun ranging from new FDA requirements for risk management plans, to creation of a new FDA Drug Safety and Risk Management Advisory Committee, to issuing risk minimization and management guidances. More information is provided in Chapters 8 and 29.

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".[132] In this framework, the concern is not about substandard or negligent care, but rather, is 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 ADEs, prolonging hospital stays by

2 days, and increase costs by $2000–2600 per patient.[133–136] Over 7000 US deaths were attributed to medication errors in 1993.[137] Although these estimates have been disputed,[138–143] the overall importance of reducing these errors has not been questioned. In recognition of this problem, 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 45.

The 1990s and especially the 2000s have seen 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, etc. These new foci are discussed in more detail in Parts IV and V of this book.

Also, 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,[144,145] there has been increased concern about reliance solely on non-experimental methods to study drug safety after marketing.[146–149] This has led to increased use of massive randomized clinical trials as part of postmarketing surveillance (see Chapter 36). 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.[150]

Finally, with the advent of the Obama administration in the US, there is 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".[151] By this definition, CER includes three key elements: (i) evidence synthesis, (ii) evidence generation, and (iii) evidence dissemination. Typically, CER is conducted through observational studies of either large administrative or medical record databases (see Part III, Section B), or large naturalistic clinical trials (see Chapter 36). In many ways, the UK has been focusing on CER for years, with 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.[152] However, the Obama administration included $1.1 billion for CER in its federal stimulus package, and has 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 compared to another. CER extends beyond pharmacoepidemiology in that CER 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 carried out in CER has been in pharmacoepidemiology. See Chapter 32 for more discussion of CER.

The current drug approval process

Drug approval in the US

Since the mid-1990s, there has been a decline in the number of novel drugs approved per year,[101,153] while the cost of bringing a drug to market has risen sharply.[154] The total cost of drug development to the pharmaceutical industry increased from $24 billion in 1999, to $32 billion in 2002,[155] and to $65.2 billion on research and development in 2008.[156] The cost to discover and develop a drug that successfully reached the market rose from over $800 million in 2004[157] to an estimated $1.3 billion to 1.7 billion currently.[158] In addition to the sizeable

costs of research and development, a substantial part of this 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 which 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 the possible efficacy of the drug. 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 a drug's efficacy and to 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 one 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 which occur in six 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 gain little additional information 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, the implementation varies widely. A WHO comparative analysis of drug regulation in ten countries found that not all countries even have a written national drug policy document.[159] Regulation of medicines in some countries is centralized in a single agency, which 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,[160] the State Food and Drug Administration in China,[161] the Medicines Agency in Denmark,[162] the Medicines Agency in Norway,[163] the Center for Drug Administration in Singapore,[164] and the Medicines and Medical Devices Safety Authority in New Zealand.[165] 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.[159] 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).[164] 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,[162] India,[166] and New Zealand[165] 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 Minister of Health.[159] 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.[159] Some countries carry out independent assessments while others, especially many developing countries, rely on WHO assessments or other sources.[159] 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 health-care professionals in reporting adverse drug reactions.[159] Another source noted that countries also differ on the extent of emphasis on quantitative or qualitative analysis for assessing pre- and postmarketing data.[167]

Further, within Europe, each country has its own regulatory agency, for example the United Kingdom's 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 has shifted away from the national licensing authorities of the EU members to that of the centralized authority of the European Medicines Evaluation Agency (EMEA), which was established in 1993.[168] 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, for example, and would compromise patient safety.[169] Now called the European Medicines Agency (EMA), the EMA is a decentralized body of the European Union, 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 European Union countries, but individual national medicines agencies are responsible for monitoring the safety of approved drugs and sharing this information with EMA.[170]

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.[171] Now, since the passage of the Food and Drug Administration Amendments Act of 2007 (FDAAA-PL 110-85) noted above, the FDA has the right to require 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

Table 1.1 Potential contributions of pharmacoepidemiology

A. Information that supplements the information available from premarketing studies—better quantitation of the incidence of known adverse and beneficial effects
   a. Higher precision
   b. In patients not studied prior to marketing, e.g., the elderly, children, pregnant women
   c. As modified by other drugs and other illnesses
   d. 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
      a. Uncommon effects
      b. 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

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 this information.

Supplementary information

Premarketing studies of drug effects are necessarily limited in size. After marketing, non-experimental 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,[172,173] but the FDA requested the manufacturer to conduct a postmarketing surveillance study of the drug in the US to quantitate its incidence more precisely.[118] In recent years, there has even been an attempt, in selected special cases, to release selected critically important drugs more quickly, by taking advantage of the work that can be performed after marketing. Probably the best-known example was zidovudine.[174,175] As noted above, the increased sample size available after marketing also permits a more precise determination of the correct dose to be used.[99,101,176,177]

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 await studies conducted after drug marketing.[178]

Additionally, for reasons of statistical efficiency, premarketing clinical trials generally seek subjects who are as homogeneous 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, etc.[179] For example, after marketing, the ophthalmic preparation of timolol was noted to cause many serious episodes of heart block and asthma, resulting in over ten deaths. These effects were not detected prior to marketing, as patients with underlying cardiovascular or respiratory disease were excluded from the premarketing studies.[180]

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 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 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 very new focus on comparative effectiveness research (see Chapter 32).

New types of information not available from premarketing studies

As mentioned above, 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.[181]

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 24 and 25).

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.[182]

Finally, it is only in the past decade or two that our society has 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.[183] It is clear that the 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.[184] Conversely, a drug's beneficial effects could reduce the need for medical care, resulting in savings that can 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 Chapters 31 and 38).

General contributions of pharmacoepidemiology

Lastly, it is important to review the general contributions that can be made by pharmacoepidemiology. 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 can be made by pharmacoepidemiologic studies. 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 which may be there. 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, England: 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 sulfanilamide (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**: 114–9.
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: Wiley, 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: 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, England, 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, p. 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, Sayler ME, Rampey AH, Heiligenstein JH, *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, Burgess C, Jackson R, Kwong T, *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, Crane J, Burgess C, Culling C, *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, Crane J, Burgess C, Keane A, *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, Horwitz RI, Habbick B, Cockcroft D, *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, Corti MC, Salive ME, Cerhan JR, *et al*. Calcium-channel blockade and incidence of cancer in aged populations. *Lancet* 1996; **348**: 493–7.
56. Braun S, Boyko V, Behar S, Reicher-Reiss H, Shotan A, Schlesinger Z, *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, Strom BL, Stolley PD, Zauber AG, *et al.* Calcium channel blockers and the risk of cancer. *JAMA* 1998; **279**: 1000–4.
61. Michels KB, Rosner BA, Walker AM, Stampfer MJ, Manson JE, Colditz GA, *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, Gurwitz J, Andrade SE, Goodman M, *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 Safety* 2001; **10**: 211–8.
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, Hensrud DD, Edwards BS, Edwards WD, *et al*. Valvular heart disease associated with fenfluramine-phentermine. *N Engl J Med* 1997; **337**: 581–8.
68. Kimmel SE, Keane MG, Crary JL, Jones J, Kinman JL, Beare J, *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, Spitz PW, Hennessy S, Feldman H, *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, Hennessy S, Kimmel SE, Farrar J, *et al.* Parenteral ketorolac: the risk for acute renal failure. *Ann Intern Med* 1997; **126**: 193–9.
71. Hennessy S, Kinman JL, Berlin JA, Feldman HI, Carson JL, Kimmel SE, *et al.* Lack of hepatotoxic effects of parenteral ketorolac in the hospital setting. *Arch Intern Med* 1997; **157**: 2510–4.
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, Switzerland: World Health Organization, 1998, ISBN 92 4 120877 5.
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, Ramrakhiani S, Quiason SG, Phillips NJ, *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. 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, Broderick JP, Brott T, Feldmann E, *et al*. Phenylpropanolamine and the risk of hemorrhagic stroke. *N Engl J Med* 2000; **343**: 1826–32.
92. Lathrop SL, Ball R, Haber P, Mootrey GT, Braun MM, Shadomy SV, *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. Committee to Assess the Safety and Efficacy of the Anthrax Vaccine. *The Anthrax Vaccine: Is it safe? Does it work?* Washington, DC: National Academy 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. FDA. *Drug Safety Labeling Changes*. http://www.fda.gov/Safety/MedWatch/SafetyInformation/Safety-RelatedDrugLabelingChanges/default.htm (accessed May 2010).
98. US General Accounting 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. GAO FDA Drug Safety Report. *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 may 2011).
101. Heerdink ER, Urquhart J, Hubert G, Leufkens HG. Changes in prescribed drug doses after market introduction. *Pharmacoepidemiol Drug Saf* 2002; **11**: 447–53.
102. Oliver K. *Drug Approval in the United States: Problems and Solutions*. http://leda.law.harvard.edu/leda/data/40/koliver.html (accessed Dec. 2010).
103. FDA. *Prescription Drug User Fee Act (PDUFA)*. http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/default.htm (accessed Nov. 2010).
104. Friedman M. *Testimony on Prescription Drug User Fee Act of 1992 (PDUFA)*. Food and Drug Administration, April 23, 1997. http://www.hhs.gov/asl/testify/t970423a.html (accessed May 2011).
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*. National Academy of Sciences, 2007.
111. PDUFA IV fees will go to improve postmarket drug safety. *FDA news Drug Daily Bulletin* 2008; **5** (89). http://www.fdanews.com/newsletter/article?articleId=106457&issueId=11559 (accessed May 2011).
112. Lee JAH, Draper PA, Weatherall M. Prescribing in three English towns. *Milbank Mem Fund Quart* 1965; **43**: 285–90.
113. Muller C. Medical review of prescribing. *J Chronic Dis* 1965; **18**: 689–96.
114. Meade TW. Prescribing of chloramphenicol in general practice. *BMJ* 1967; **1**: 671–4.

115. 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.
116. Muller C. Outpatient drug prescribing related to clinic utilization in four New York City hospitals. *Health Serv Res* 1968; **3**: 142–54.
117. 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*. Massachusetts: PSG, 1977, pp. 59–82.
118. Joint Commission on Prescription Drug Use. *Final Report*. Washington, 1980.
119. Strom BL, Carson JL, Morse ML, Leroy AA. The Computerized Online Medicaid Analysis and Surveillance System: a new resource for postmarketing drug surveillance. *Clin Pharmacol Ther* 1985; **38**: 359–64.
120. Inman WHW. Prescription event monitoring. *Acta Med Scand Suppl* 1984; **683**: 119–26.
121. 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.
122. Andrews EA, Avorn J, Bortnichak EA, Chen R, Dai WS, Dieck GS, *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.
123. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**: 200–8.
124. Neutel CI. Privacy issues in research using record linkage. *Pharmacoepidemiol Drug Saf* 1997; **6**: 367–9.
125. Melton LJ III. The threat to medical-records research. *N Engl J Med* 1997; **337**: 1466–70.
126. Mann RD. Data privacy and confidentiality. *Pharmacoepidemiol Drug Saf* 1999; **8**: 245.
127. Andrews EB. Data privacy, medical record confidentiality, and research in the interest of public health. *Pharmacoepidemiol Drug Saf* 1999; **8**: 247–60.
128. Mann RD. The issue of data privacy and confidentiality in Europe–1998. *Pharmacoepidemiol Drug Saf* 1999; **8**: 261–4.
129. Califf RM. The need for a national infrastructure to improve the rational use of therapeutics. *Pharmacoepidemiol Drug Saf* 2002; **11**: 319–27.
130. *Public Law 108-173. Medicare Prescription Drug, Improvement, and Modernization Act of 2003. Sec. 1013. Research on Outcomes of Health Care Items and Services*. http://thomas.loc.gov/cgi-bin/bdquery/z?d108:HR00001:@@@D&summ2=m& (accessed May 2011).
131. *AHRQ Launches New "Effective Health Care Program" To Compare Medical Treatments and Help Put Proven Treatments into Practice*. Press Release Date: September 29, 2005. http://www.ahrq.gov/news/press/pr2005/effectivepr.htm (accessed May 2011).
132. Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err Is Human. Building a Safer Health System*. Washington, DC: National Academy Press, 2000.
133. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, *et al.* Incidence of adverse drug events and potential adverse drug events: implications for prevention. *JAMA* 1995; **274**: 29–34.
134. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, *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.
135. 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.
136. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, *et al*. The costs of adverse drug events in hospitalized patients. *JAMA* 1997; **277**: 307–11.
137. Phillips DP, Christenfeld N, Glynn LM. Increase in US medication-error deaths between 1983 and 1993. *Lancet* 1998; **351**: 643–4.
138. Manasse HR. Increase in US medication-error deaths (letter). *Lancet* 1998; **351**: 1655.
139. Ferner RE, Anton C. Increase in US medication-error deaths (letter). *Lancet* 1998; **351**: 1655–6.
140. Rooney C. Increase in US medication-error deaths (letter). *Lancet* 1998; **351**: 1656–7.
141. Phillips DP, Christenfeld N, Glynn LM. Increase in US medication-error deaths (letter). *Lancet* 1998; **351**: 1657.
142. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in the Institute of Medicine report. *JAMA* 2000; **284**: 93–5.
143. Leape LL. Institute of Medicine error figures are not exaggerated. *JAMA* 2000; **284**: 95–7.
144. Manson JE, Hsia J, Johnson KC, Rossouw JE, Assaf AR, Lasser NL, *et al.* Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med* 2003; **349**: 523–34.
145. 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.
146. 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.
147. Grimes DA, Lobo RA. Perspectives on the Women's Health Initiative trial of hormone replacement therapy. *Obstet Gynecol* 2002; **100**: 1344–53.
148. Whittemore AS, McGuire V. Observational studies and randomized trials of hormone replacement therapy: what can we learn from them? *Epidemiology* 2003; **14**: 8–9.
149. Piantadosi S. Larger lessons from the Women's Health Initiative. *Epidemiology* 2003; **14**: 6–7.
150. *GAO Report (GAO-09-866). New Drug Approval: FDA needs to Enhance its Oversight of Drugs Approved on the Basis of Surrogate Endpoints*. September 23, 2009. http://www.gao.gov/highlights/d09866high.pdf (accessed May 2011).
151. Federal Coordinating Council for Comparative Effectiveness Research. *Report to the President and Congress on comparative effectiveness research*. Washington, DC, 2009.
152. National Institute for Health and Clinical Excellence. *What We Do*. Updated 2010. http://www.nice.org.uk/aboutnice/whatwedo/what_we_do.jsp (accessed May 2011).
153. Zhe-Jie Ni. *New drugs approved by FDA in 2004–2005*. Trends in Bio/Pharmaceutical Industry, 2006. http://tbiweb.org/tbi/file_dir/TBI2006/TBI2006_1_3.pdf (accessed May 2011).
154. Allan B. Haberman, Overcoming Phase II attrition problem. *Genetic Engineering and Biotechnology News* 2009; **29**, 1.
155. Pharmaceutical Research and Manufacturers of America (PhRMA). *Pharmaceutical Industry Profile 2003*. Washington, DC: PhRMA, 2003, p. 75.
156. Pharmaceutical Research and Manufacturers of America (PhRMA). *Pharmaceutical Industry Profile 2009*. Washington, DC: PhRMA, 2003, p. 2.
157. Cost to bring drug to market: $802m. http://healthcare-economist.com/2006/04/29/802m/ (accessed May 2011).
158. Collier R. Drug development cost estimates hard to swallow. *CMAJ* 2009; **180**: 279–80.
159. Ratanawijitrasin S, Wondemagegnehu E. *Effective Drug Regulation: A Multicountry Study*. WHO report, 2002. http://archives.who.int/tbs/qual/s2300e.pdf (accessed May 2011).
160. Health Canada. http://www.hc-sc.gc.ca/index-eng.php (accessed May 2011).
161. The State Food & Drug Administration in China. http://www.bing.com/search?srch=105&FORM=AS5&q=State+Food+%26+Drug+Administration+in+china (accessed May 2011).
162. Medicines Agency in Denmark. http://www.dkma.dk (accessed May 2011).
163. Medicines Agency in Norway. http://www.legemiddelverket.net/templates/InterPage_16645.aspx (accessed May 2011).
164. *Singapore's Pharmaceutical and Medical Device Regulatory Environment: 2005 Update*. http://www.pacificbridgemedical.com/publications/singapore/2005_singapores_pharmaceutical_and (accessed May 2011).
165. Medicines & Medical Devices Safety Authority in New Zealand. http://www.medsafe.govt.nz/other/about.asp (accessed May 2011).
166. Central Drugs Standard Control Organization in India. http://www.cdsco.nic.in/ (accessed May 2011).
167. Reichert JM. Pharmacopolitics: Drug regulation in the United States and Germany. *N Engl J Med* 2004; **350**: 2220–1.
168. European Medicines Agency. *About Us*. http://www.ema.europa.eu/ema/index.jsp?curl=pages/about_us/general/general_content_000235.jsp&murl=menus/about_us/about_us.jsp&mid=&jsenabled=true (accessed May 2011).
169. 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.
170. European Medicines Agency. *What Do We Do?* http://www.ema.europa.eu/ema/index.jsp?curl=pages/about_us/general/general_content_000091.jsp&murl=menus/about_us/about_us.jsp&mid=WC0b01ac0580028a42 (accessed May 2011).
171. Department of Health and Human Services. Office of Inspector General. *Postmarketing Studies of Prescription Drugs*, May 1996, OEI-03-94-00760, page 11. http://oig.hhs.gov/oei/reports/oei-03-94-00760.pdf (accessed May 2011).
172. Rosendorff C. Prazosin: severe side effects are dose-dependent. *BMJ* 1976; **2**: 508.
173. Graham RM, Thornell IR, Gain JM, Bagnoli C, Oates HF, Stokes GS. Prazosin: the first-dose phenomenon. *BMJ* 1976; **2**: 1293–4.
174. Kaitin KI. Case studies of expedited review: AZT and L-Dopa. *Law Med Health Care* 1991; **19**: 242–6.

175. Wastila LJ, Lasagna L. The history of zidovudine (AZT). *J Clin Res Pharmacoepidemiol* 1990; **4**: 25–37.
176. Peck CC. Postmarketing drug dosage changes. *Pharmacoepidemiol Drug Saf* 2003; **12**: 425–6.
177. Temple RJ. Defining dose decrease. *Pharmacoepidemiol Drug Saf* 2003; **12**: 151–2.
178. McKenzie MW, Marchall GL, Netzloff ML, Cluff LE. Adverse drug reactions leading to hospitalization in children. *J Pediatr* 1976; **89**: 487–90.
179. May FE, Stewart RB, Cluff LE. Drug interactions and multiple drug administration. *Clin Pharmacol Ther* 1977; **22**: 322–8.
180. 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.
181. 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.
182. 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.
183. 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.
184. 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.

CHAPTER 2
Basic Principles of Clinical Pharmacology Relevant to Pharmacoepidemiologic Studies

Jeffrey S. Barrett and Athena F. Zuppa
Laboratory for Applied Pharmacokinetics and Pharmacodynamics, Division of Clinical Pharmacology and Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

Introduction

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, particularly their therapeutic value in humans, including their toxicology, safety, pharmacodynamics, and pharmacokinetics. While the foundation of the discipline is underpinned by basic pharmacology (the study of the interactions that occur between a living organism and exogenous chemicals that alter normal biochemical function), the important emphasis of clinical pharmacology 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. The main objective is to promote the safe and effective use of drugs, maximizing the beneficial drug effects while minimizing harmful side effects. It is important that caregivers are skilled in the areas of drug information, medication safety, and other aspects of pharmacy practice related to clinical pharmacology. Clinical pharmacology is an important bridging discipline that includes knowledge about the relationships between: dose and exposure at the site of action (pharmacokinetics); exposure at the site of action and clinical response (pharmacodynamics); and between clinical response and outcomes. In the process, it defines the therapeutic window (the dosage of a medication between the minimum amount that gives a desired effect and the minimum amount that gives more adverse effects than desired effects) of a drug in various patient populations. Likewise, clinical pharmacology also guides dose modifications in various patient subpopulations (e.g., pediatrics, pregnancy, elderly, and organ impairment) and/or dose adjustments for various lifestyle factors (e.g., food, time of day, drug interactions).

The discovery and development of new medicines is reliant upon clinical pharmacologic 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 pharmacologic 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 health-care providers.

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

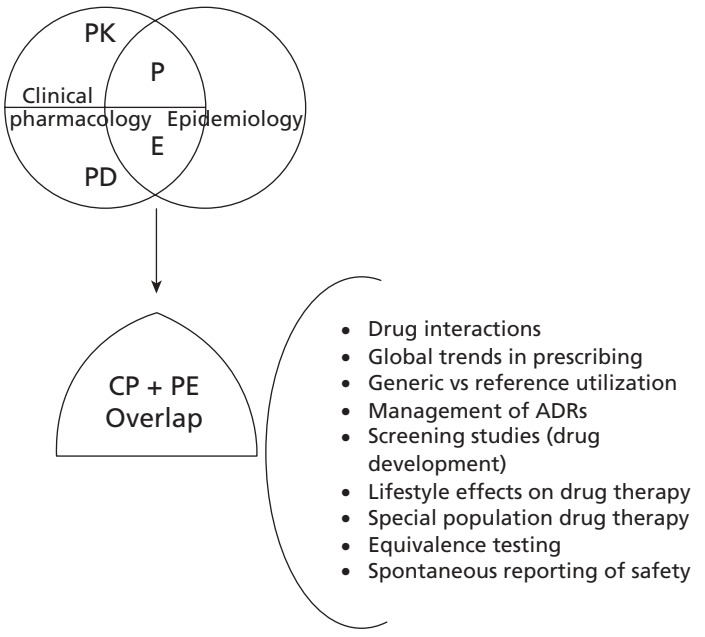
Clinical pharmacology and pharmacoepidemiology

Pharmacoepidemiology is the study of the use and effects of drugs in large numbers of people.

Studies that estimate the probability and magnitude of beneficial effects in populations, or the probability and magnitude 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 properties, interactions, toxicology, and effects (both desirable and undesirable) that can be used in therapy of diseases. As described above, underlying the discipline of clinical pharmacology are the fields of pharmacokinetics and pharmacodynamics, and each of these disciplines can be further defined by the underlying processes which dictate specific pathways (e.g., absorption, distribution, metabolism, elimination). Clinical pharmacology is essential to both our understanding of how drugs work as well as how to guide their administration. Individual pharmacotherapy can be challenging because of physiologic factors that may alter drug kinetics (e.g., age, size, etc.), pathophysiologic differences that may alter pharmacodynamics, disease etiologies in studied patients that may differ from those present in the general population, and other factors that may result in great variation in safety and efficacy outcomes. The challenge becomes more difficult when one considers critically ill populations and the paucity of well-controlled clinical trials in vulnerable populations. Likewise, prescribing caregivers of critically ill and other difficult to manage patients must have some understanding of the practices that govern



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. PK, pharmacokinetics; PD, pharmacodynamics; PE, pharmacoepidemiology; CP, clinical pharmacology.

the current dosing recommendations for their patients.

Pharmacokinetics

Pharmacokinetics refers to the study of the 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 below in greater detail.

Absorption

Absorption is the process of drug transfer from its site of administration to the blood stream. 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 intravenous, the bioavailability is usually less. Bioavailability is reduced by incomplete absorption, first-pass metabolism (defined below), and distribution into other tissues.

Volume of distribution

The *apparent volume of distribution* (V_d) 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 if it were only fluid 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 drug (C) in the blood or plasma: $V_d = \text{Drug}/C$.

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

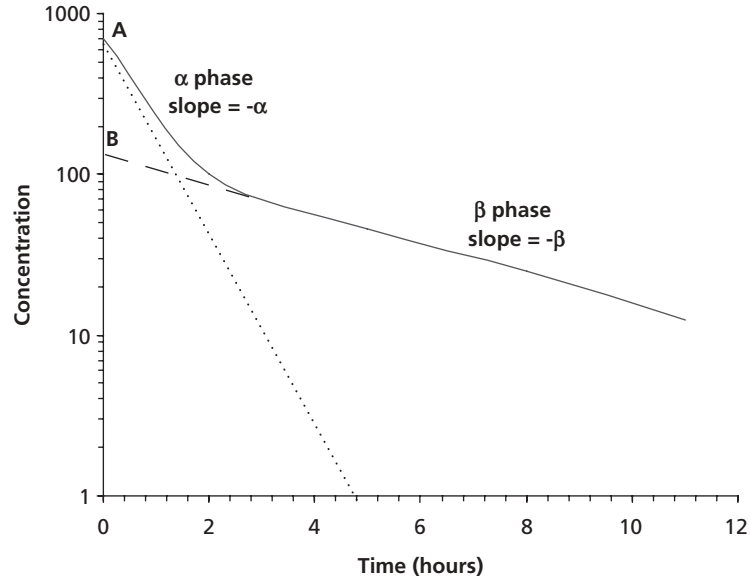


Figure 2.2 Semilogarithmic plot of concentration versus time after an intravenous administration of a drug that follows two-compartment pharmacokinetics.

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 vertical axis. The extrapolation of the β phase defines B. The dotted line is generated by subtracting the extrapolated line from the original concentration line. This second line defines α and A. The plasma concentration can be estimated 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 α and β phase may be subtle and essentially a single phase best describes the decline in drug concentration.

Metabolism

The *metabolism* of drugs is catalyzed by enzymes, and most reactions follow Michaelis–Menten kinetics: V (rate of drug metabolism) = $[((V_{max})(C)/K_m) + (C)]$, where C is the drug concentration (expressed in units such as μM), V_{max} is the maximum rate of metabolism in units of amount of product over time, typically μmol/min, and K_m is the Michaelis–Menten constant (equivalent to the substrate concentration at which the rate of conversion is half of V_{max}) also in units of concentration.[1] In most situations, the drug concentration is much less than K_m and the equation simplifies to: $V = (V_{max})(C)/K_m$. In this case, the rate of drug metabolism is directly proportional to the concentration of free drug, and follows first-order kinetic theory. A constant percentage of the drug is metabolized per unit time, and the absolute amount of drug eliminated per unit time 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 K_m, and the Michaelis–Menten equation reduces to: V (rate of drug metabolism) = $(V_{max})(C)/(C) = V_{max}$. 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 time. For drugs that follow zero-order elimination, a large increase in serum concentration can result from a small increase in dose.

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 greater 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 (see Chapter 34), 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, 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 an ability to produce reactive oxygen radicals through a reduction of O_2 and is the only CYP system strongly induced (drug molecule initiates or enhances the expression of an enzyme) by alcohol, which is itself a substrate (a molecule upon which an enzyme acts). 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 non-specific 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 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 non-renal 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 of the nephron, where active transport systems secrete primarily organic acids and bases. Organic acids include most cephalosporins, loop diuretics, methotrexate, non-steroidal anti-inflammatories, penicillins, and thiazide diuretics. Organic bases include ranitidine and morphine. As drugs move toward the distal convoluting tubule, the concentration increases. High urine flow rates decrease the concentration of drug in the distal tubule, decreasing the likelihood that a drug will diffuse from the lumen. For both weak acids and bases, the non-ionized 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 alkalinization 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 V_d, and inversely proportional to Cl[1]: $t_{1/2\beta} = (0.693)(V_d)/Cl$.

Special populations

The term "special populations" as applied to drug development refers to discussions in the early 1990s involving 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, 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 pharmacologic 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 physiological 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 an elderly and young population. Table 2.1 lists the primary physiologic factors affected by aging.[7] Recognition of these factors is important for predicting the implications of aging on 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 significant first-pass metabolism 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., the ACE (angiotensin converting enzyme) inhibitors enalapril and perindopril) are likely to experience reduction in this phase and therefore reduced exposure to the active species.

Based on age-related changes in body composition, polar drugs that 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, non-polar compounds are often lipid soluble and exhibit larger volumes of distribution in older patients. The impact of the larger V_d is

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 number of 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

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

Several drug classes, including water-soluble antibiotics, diuretics, water-soluble beta-adrenoceptor blockers, and non-steroidal anti-inflammatory drugs,[7,15] exhibit changes in clearance with age because 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 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 target an elderly subpopulation in a separate trial, consistent with FDA guidance. The FDA guideline for studies in the elderly is directed principally toward new molecular entities likely to have significant use in the elderly, 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 affect drug disposition.[19,20]

Renal. Renal function in the premature and full-term neonate, both glomerular filtration and tubular secretion, is significantly reduced, as

compared to older children. Maturation of renal function is a dynamic process which 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 to 4 mL/min/1.73 m^2 in full-term neonates, but it may be as low as 0.6 to 0.8 mL/min/1.73 m^2 in preterm neonates. The glomerular filtration rate increases rapidly during the first 2 weeks of life and continues to rise until adult values are reached at 8 to 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% 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 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 to 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 glucuronosyl transferase (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 as 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 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 impact 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 percent 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%, as 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,

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

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 (normal pregnancy, labor, and delivery). Few well-controlled studies of therapeutic drugs 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) (Table 2.2).

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 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 with pregnancy. Likewise, the metabolism of drugs via select P450-mediated pathways (3A4, 2D6, and 2C9) and UGT isoenzymes are 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; more clinical studies to determine the effect of pregnancy on the pharmacokinetics and pharmacodynamics of drugs commonly used in pregnancy are sorely needed. (See Chapter 28 for studies of drug-induced birth defects.)

Organ impairment

Renal dysfunction. Renal failure can influence 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 V_d.[15] Drugs that are more than 30% eliminated unchanged in the urine are likely to have significantly diminished *Cl* 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

normally are highly protein-bound. The presence of significant edema and ascites may alter the V_d 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 V_d. 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 (often many) drug, increasing the chance of a drug–drug interaction. Pharmacokinetic 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 metabolism of other drugs and lower plasma concentrations. Common inducers include barbiturates, carbamezapine, 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 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 (PD), in general terms, seeks to define what the drug does to the body (i.e., the effects or response to drug therapy). Pharmacodynamic modeling attempts to characterize measured, physiological parameters before and after drug administration, with the effect defined as the change in a physiological parameter relative to its predose or baseline value. Baseline is defined as the physiological 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. In this case, we refer to clinical and not necessarily economic benefits, though there clearly may be concordance. 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, that is the concentration–effect relationship, 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 among drug entities in the pharmacodynamic time course can be considered as being direct or indirect. A direct effect is directly proportional to concentration at the site of measurement, usually the plasma. An indirect effect exhibits some type of temporal delay, either because of differences between site of action and measurement or because the effect results only after other physiologic or pharmacologic conditions are satisfied.

Direct effect relationships are easily observed with some 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.

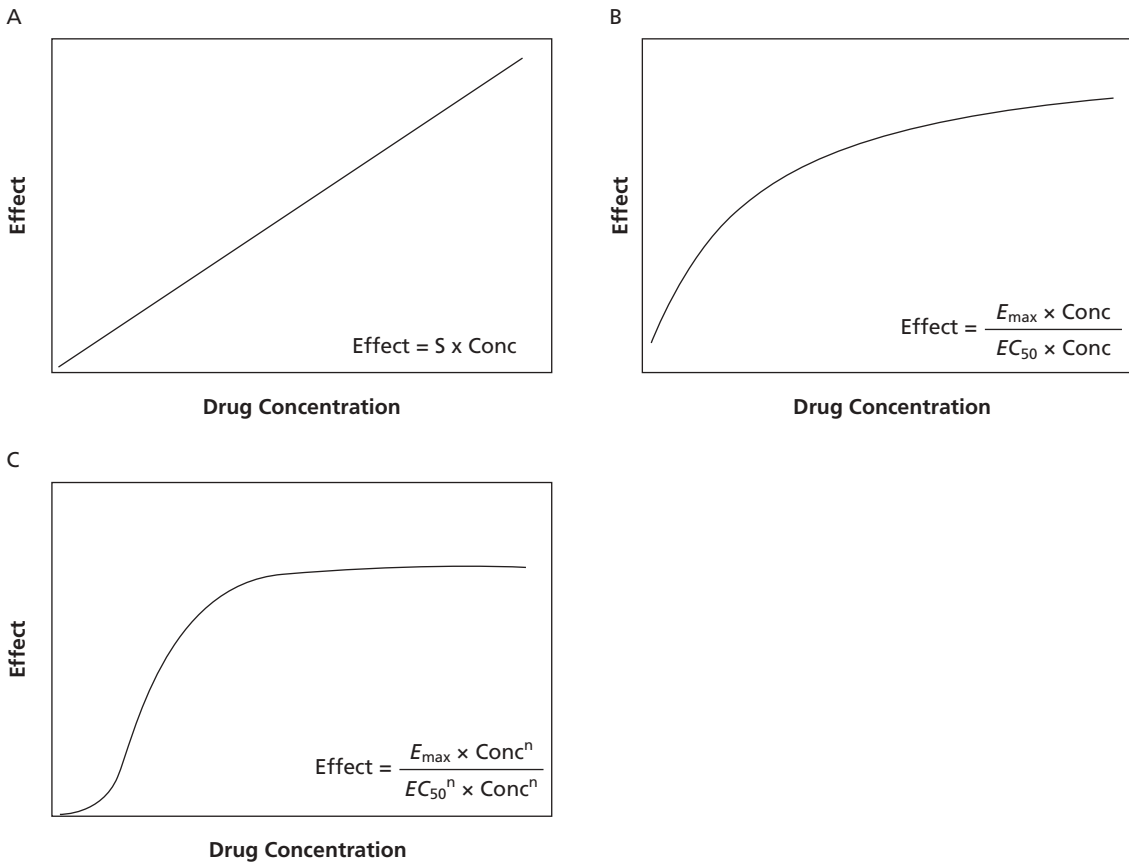


Figure 2.3 Representative pharmacodynamic relationships for drugs which exhibit direct responses: (A) linear, (B) hyperbolic, and (C) sigmoid–E_{max} relationships shown. S is the slope of the linear response; E_{max} refers to the maximum effect observed; EC_{50} refers to the concentration at which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

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. Hysteresis refers to the phenomenon where there is a time-lapse between the 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 (e.g., thiopental, fentanyl, many others). Three conditions are predominantly responsible for hysteresis: the biophase (actual site of drug action) is not in the central compartment (i.e., plasma or blood 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 concentrations (C_e) are used as opposed to the plasma concentrations (C_p). In this situation, a hysteresis loop is observed when plotting C_e versus C_p (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]

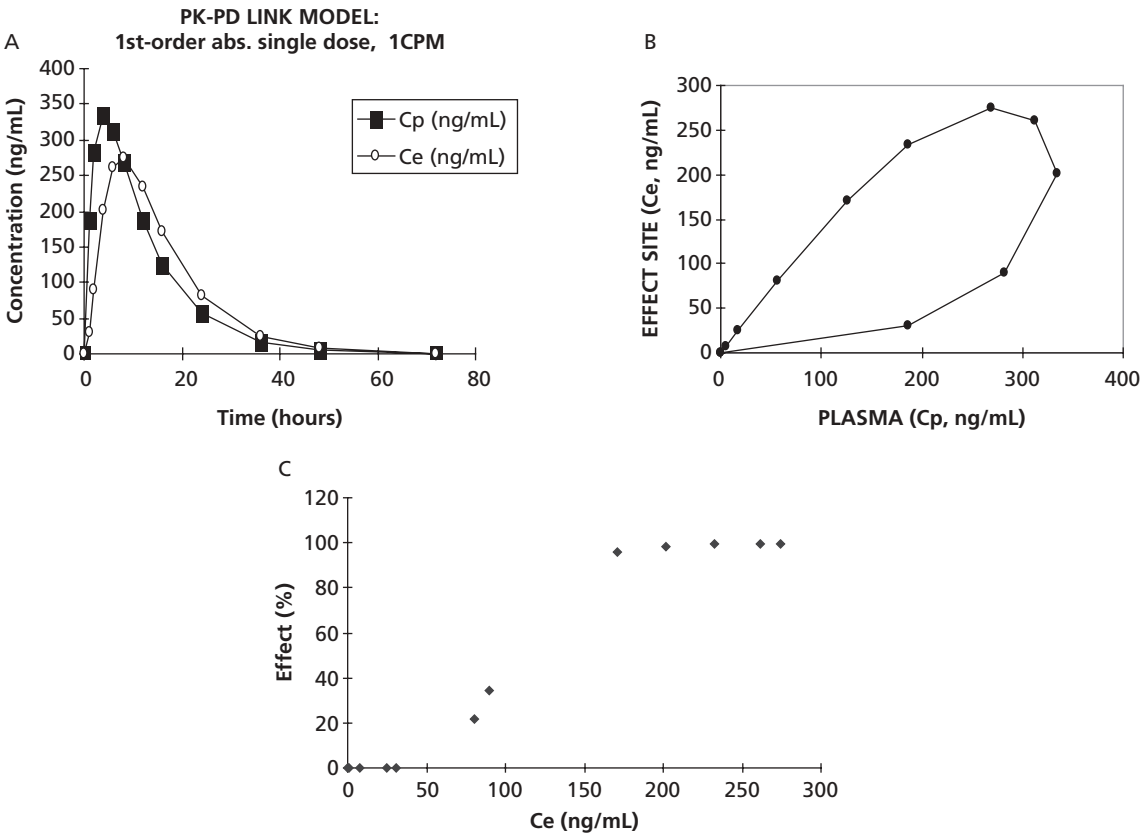


Figure 2.4 (A) Concentration–time, (B) hysteresis, and (C) effect–concentration plots (clockwise order) illustrating the use of an effect compartment to explain observed hysteresis. Cp, plasma concentration; Ce, concentration at the effect site; 1CPM, one compartment model.

The salient point is that pharmacodynamic characterization, and likewise dosing guidance derived from such investigation, stands 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 older children and adults with respect to immunosuppressive effects of cyclosporine,[39] and 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. 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 comprehensive knowledge of genes, proteins, and single nucleotide polymorphisms. Genetic variations, or SNPs (single nucleotide polymorphisms), 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; 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 single nucleotide polymorphisms. One study found that at 3, 6, and 12 months post heart transplantation, there was a significant difference in tacrolimus blood concentrations per dose/kg/day between the CYP3A5 *1/*3 (CYP3A5 expressor) and the *3/*3 (non-expressor) 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 34 for Molecular Pharmacoepidemiology.)

Conclusion

Clinical pharmacology serves an important role in the development of new drugs and the management of pharmacotherapy. It provides essential knowledge needed to inform the drug developer, the investigator or trialist, the regulator, and the caregiver, each 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 impact (clinical and economic) of a new drug once available to the marketplace can often be forecast based on how the agent's clinical pharmacologic attributes compare to existing therapies. Likewise, the connection between utilization, compliance, 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 drugs behave in humans. In an era when more holistic approaches for the care of patients are sought to maintain a healthy overall well-being and avoid chronic and severe disease, clinical strategies are

likely to engage more preventative approaches. Likewise, clinical pharmacology and pharmacoepidemiology will be essential disciplines that discriminate strategies (holistic and preventative) that are truly beneficial from those that are not, or are even harmful.

References

1. Wagner JG. *Biopharmaceutics and Relevant Pharmacokinetics*. Hamilton: Drug Intelligence Pubns, 1971.
2. Woolf TF, ed. *Handbook of Drug Metabolism*. New York: Marcel Dekker, 1999.
3. Chen W, Koenigs LL, Thompson SJ, Peter RM, Rettie AE, Trager WF, *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**: 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**: 275–82.
5. Kostrubsky, SE, Sinclair JF, Strom SC, Wood S, Urda E, Stolz DB, *et al*. Phenobarbital and phenytoin increased acetaminophen hepatotoxicity due to inhibition of UDP-glucuronosyltransferases in cultured human hepatocytes. *Toxicol Sci* 2005; **87**: 146–55.
6. Gibaldi M, Perrier D. *Pharmacokinetics*, 2nd edn. New York: Marcel Dekker, 1982.
7. Mangoni AA, Jackson SH. Age-related changes in pharmacokinetics and pharmacodynamics: basic principles and practical applications. *Br J Clin Pharmacol* 2004; **57**: 6–14.
8. Montgomery RD, Haeney MR, Ross IN, Sammons HG, Barford AV, Balakrishnan S, *et al*. The ageing gut: a study of intestinal absorption in relation to nutrition in the elderly. *Q J Med* 1978; **47**: 197–24.
9. Saltzman JR, Russell RM. The aging gut. Nutritional issues. *Gastroenterol Clin North Am* 1998; **27**: 309–24.
10. Bhutto A, Morley JE. The clinical significance of gastrointestinal changes with aging. *Curr Opin Clin Nutr Metab Care* 2008; **11**: 651–60.
11. Stanski DR, Maitre PO. Population pharmacokinetics and pharmacodynamics of thiopental: the effect of age revisited. *Anesthesiology* 1990; **72**: 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**: 395–402.
13. Karim A, Noveck R, McMahon FG, Smith M, Crosby S, Adams M, *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**: 267–78.
14. Wallace SM, Verbeeck RK. Plasma protein binding of drugs in the elderly. *Clin Pharmacokinet* 1987; **12**: 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**: 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**: 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**: 2667–73.
18. Durnas C, Loi CM, Cusack BJ. Hepatic drug metabolism and aging. *Clin Pharmacokinet* 1990; **19**: 359–89.
19. Strolin Benedetti M, Baltes EL. Drug metabolism and disposition in children. *Fundam Clin Pharmacol* 2003; **17**: 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**: E475–87.
21. Johnson TN. The problems in scaling adult drug doses to children. *Arch Dis Child* 2008; **93**: 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**: 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**: 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**: 411–15.
25. Knibbe CA, Krekels EH, van den Anker JN, DeJongh J, Santen GW, van Dijk M, *et al*. Morphine glucuronidation in preterm neonates, infants and children

younger than 3 years. *Clin Pharmacokinet* 2009; **48**: 371–85.
26. Riezzo G, Indrio F, Raimondi F, Montagna O, Salvia G, Massimo B, *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**: 6.
27. Lebenthal A, Lebenthal E. The ontogeny of the small intestinal epithelium. *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**: 255–8.
29. Aranda JV, Varvarigou A, Beharry K, Bansal R, Bardin C, Modanlou H, *et al*. Pharmacokinetics and protein binding of intravenous ibuprofen in the premature newborn infant. *Acta Paediatr* 1997; **86**: 289–93.
30. Onishi S, Ohki Y, Nishimura Y, Itoh S, Isobe K, Hosoe A, *et al*. Distribution of phenobarbital in serum, brain and other organs from pediatric patients. *Dev Pharmacol Ther* 1984; **7**: 153–9.
31. Dawes M, Chowienczyk PJ. Drugs in pregnancy. Pharmacokinetics in pregnancy. *Best Pract Res Clin Obstet Gynaecol* 2001; **15**: 819–26.
32. Anderson GD. Pregnancy-induced changes in pharmacokinetics: a mechanistic-based approach. *Clin Pharmacokinet* 2005; **44**: 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**: 1841–7.
34. Shammas FV, Dickstein K. Clinical pharmacokinetics in heart failure. An updated review. *Clin Pharmacokinet* 1988; **15**: 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**: 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**: 414–20.
37. Bruguerolle B. Chronopharmacokinetics. Current status. *Clin Pharmacokinet* 1998; **35**: 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: Kluwer Academic/ Plenum Publishers, 2004, pp. 333–72.
39. Marshall JD, Kearns GL. Developmental pharmacodynamics of cyclosporine. *Clin Pharmacol Ther* 1999; **66**: 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**: 680–7.
41. De Cicco M, Macor F, Robieux I, Zanette G, Fantin D, Fabiani F, *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**: 332–9.

CHAPTER 3
Basic Principles of Clinical Epidemiology Relevant to Pharmacoepidemiologic Studies

Brian L. Strom
Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Pharmacoepidemiology applies the methods of epidemiology to the content area of clinical pharmacology. Chapter 2 reviewed basic principles of clinical pharmacology, the content area of pharmacoepidemiology. Therefore, in order to understand the approaches and methodologic issues specific to the field of pharmacoepidemiology, the basic principles of the field of epidemiology must be understood as well. To this 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 is 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 of 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 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 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 6 weeks later. One might expect to see the blood pressure of the 20 men treated with the active drug decrease more than the blood pressure of the 20

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

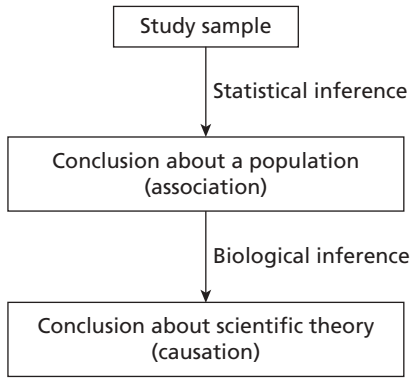


Figure 3.1 Overview of the scientific method.

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 biological 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 biological 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 below. First, however, we will place causal associations into a proper perspective, by describing the different types of errors that can be made in performing a study and the different types of associations that each results in.

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

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)

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

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]

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 47 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 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 37 and 47 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 (biological 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 above, 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 makes one 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, etc. 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 that of *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 he or she were currently taking diazepam and also whether he or she 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. A *dose–response relationship* 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 a dose–response relationship or a duration–response 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 to those at the top of the table, 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 below. This is the other major defense used by the tobacco industry. Of course, it would

Table 3.4 Advantages and disadvantages of epidemiologic study designs

Study design	Advantages	Disadvantages
Randomized clinical trial (Experimental study)	Most convincing design Only design that 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 hypothesis	Cannot be used for testing hypotheses

not be ethical or logistically feasible to randomly allocate individuals to smoke or not to smoke and expect them to be followed 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 33 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 upwards.

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 his or her 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 he or she were given another dose of naloxone the patient 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–re-challenge 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 33.

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 might also 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 "ecological 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 are studies that 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 eight 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 re-creating 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 are studies that 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.

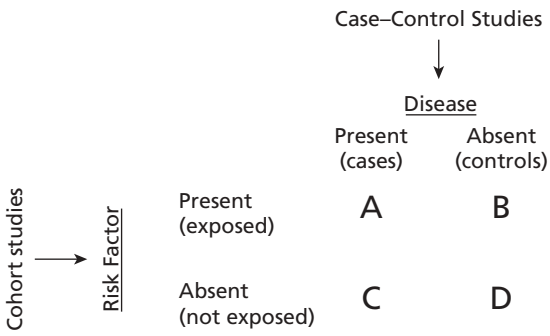


Figure 3.2 Cohort and case–control studies provide similar information, but approach data collection from opposite directions. (Reproduced from Strom BL. Medical databases in post-marketing drug surveillance. *Trends in Pharmacological Sciences* 1986; **7**: 377–80, with permission from Elsevier).

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 particularly 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.

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 *relative risk* 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. 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 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 that the diseased cases and undiseased controls were drawn from. 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 tenfold 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 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 contraceptive, 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 above.

However, even randomized clinical trials are not without their problems. The randomized clinical trial outlined above, 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 non-experimental methods to study drug safety after marketing,[45–47] and we now see increasing use of massive randomized clinical trials as part of post-marketing surveillance (see Chapter 36).

Discussion

Thus, a series of different study designs are available (Table 3.4), each with 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 *non-experimental study designs*, in order to differentiate them from experimental studies. In non-experimental 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 re-creates events that happened in the past using medical records, questionnaires, or interviews. Data can also be collected

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
      a. Experimental studies (clinical trials, intervention studies)
B. Classified by how data are collected for the study
   1. Retrospective (historical, non-concurrent, retrolective) studies
   2. Prospective (prolective) studies
   3. Cross-sectional studies

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 the terms used by a majority of epidemiologists. Unfortunately, however, other terms have been used for most of these study designs, as well. Table 3.5 also 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 used the term "retrospective study" to refer to a case–control study and used the term "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 non-experimental study, while others appear to use the term 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. What is more important than the terminology, however, are the concepts underlying the terms. Understanding these concepts, the reader can choose to use whatever terminology he or she is comfortable with.

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:

> 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.
>
> Sir Austin Bradford Hill[27]

References

1. Lilienfeld DE, Stolley P. *Foundations of Epidemiology*, 3rd edn. New York: Oxford University Press, 1994.
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, 2008.
7. Weiss N. *Clinical Epidemiology: The Study of the Outcome of Illness*, 2nd edn. New York: Oxford University Press, 1996.
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. *Clinical Epidemiology: The Essentials*, 4th edn. Lippincott Williams & Wilkins, 2005.
11. Gordis L. *Epidemiology*, 4th edn. Philadelphia, PA: Saunders, 2009.
12. Rothman KJ. *Epidemiology: An Introduction*. New York: Oxford University Press, 2002.
13. Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. Sudbury, MA: Jones and Bartlett, 2006.
14. Jekel JF, Elmore JG, Katz DL. *Epidemiology, Biostatistics, and Preventive Medicine*, 3rd edn. Philadelphia, PA: Saunders, 2007.
15. Aschengrau A, Seage GR. *Essentials of Epidemiology in Public Health*. Sudbury, MA: Jones and Bartlett, 2008.
16. Friis RH, Sellers TA. *Epidemiology for Public Health Practice*, 4th edn. Sudbury, MA: Jones & Bartlett Publishers, 2008.
17. Hulley SB, Cummings SR, Browner WS, Grady D, Newman TB. *Designing Clinical Research: An Epidemiologic Approach*, 3rd edn. Baltimore, MD: Lippincott Williams and Wilkins, 2006.
18. Greenberg RS, Daniels SR, Flanders WD, Boring JR, Eley, JW. *Medical Epidemiology*, 4th edn. USA: McGraw-Hill, 2005.
19. Wassertheil-Smoller S. *Biostatistics and Epidemiology: a Primer for Health Professionals*, 3rd edn. New York: Springer-Verlag, 2004.
20. Katz DL. *Clinical Epidemiology and Evidence-Based Medicine: Fundamental Principles of Clinical Reasoning and Research*. Thousand Oaks, CA: Sage Publications, 2001.
21. Bhopal RS. *Concepts of Epidemiology: Integrating the Ideas, Theories, Principles and Methods of Epidemiology*, 2nd edn. New York: Oxford University Press, 2008.
22. Haynes RB, Sackett DL, Guyatt GH, Tugwell P. *Clinical Epidemiology: How To Do Clinical Practice Research*. Philadelphia, PA: Lippincott-Raven Publishers, 2006.
23. Bassetti WHC, Woodward M. *Epidemiology: Study Design and Data Analysis*, 2nd edn. Boca Raton, Florida: Chapman & Hall/CRC, 2004.
24. Weiss NS, Koepsall T, Koepsell TD. *Epidemiologic Methods: Studying the Occurrence of Illness*. New York: Oxford University Press, 2004.
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, Aeugle ME, Josie ME, Wood SL, *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, p. 1–17.
38. Royal College of General Practitioners. *Oral Contraceptives and Health*. Chapter 7. London: Pitman Publishing, 1974.
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, Rossouw JE, Assaf AR, Lasser NL, *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.

CHAPTER 4
Sample Size Considerations for Pharmacoepidemiologic Studies

Brian L. Strom
Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

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 one and six in a thousand 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.

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]

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

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

For cohort studies	For case–control studies
1. α, or Type I error considered tolerable, and whether it is one-tailed or two-tailed	1. α, or Type I error considered tolerable, and whether it is one-tailed or two tailed
2. β, or Type II error considered tolerable	2. β, 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 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 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 α, and the larger the sample size that would be required. Conventionally, the α is set at 0.05, although this certainly does not have to be the case. Note that α 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 α 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 α to be two-tailed. To decide whether α should be one-tailed or two-tailed, an investigator should consider what his or her 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 usually, although not always, used.

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 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 β is the *power* of a study, that is the probability of detecting a difference if a difference really exists. Power is calculated as $(1 - \beta)$. Again, the more tolerant one is willing to be of Type II errors, that is the higher the β, the smaller the sample size required. The β 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. β 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 which 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 outcome variable of interest in the unexposed control group. Again, the more you ask of a study, 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 controls as exposed subjects. 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 greater than three or four controls per study subject. For example, one could design a study with an α 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 was discussed and illustrated above) 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 assure 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 above 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 the most often used is modified from Schlesselman:[3]

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

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 α. Note that K does not need to be an integer.

A series of tables are presented in the Appendix, which were calculated using this formula. In Tables A1 through A4 we have assumed an α (two-tailed) of 0.05, a β of 0.1 (90% power), and control to exposed ratios of 1:1, 2:1, 3:1, and 4:1, respectively. Tables A5 through A8 are similar, except they assume a β 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 table) plus the controls.

For example, what if one wanted to investigate a new non-steroidal 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 α, β, 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 α (two-tailed) of 0.05 and a β of 0.1, one could look in Table A1 and see that it would require 31483 exposed subjects, as well as an equal number of unexposed controls. If one were less concerned with missing a real finding, even if it was there, one could change β to 0.2, and the required sample size would drop to 23518 (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 13402, with four times as many controls, a total of 67010 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 α, resulting in a somewhat lower sample size of 10728, 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 α of 0.05, a β of 0.1, and one control per 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 315268 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 α, or the Type I error one is willing to tolerate; β, 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 discussed in the section on cohort studies, above.

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

Table 4.2 Examples of sample sizes needed for a cohort study

Disease incidence rate assumed in unexposed	α	β	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	31483	31483
0.001	0.05 (2-tailed)	0.2	2	1	23518	23518
0.001	0.05 (2-tailed)	0.2	2	4	13402	53608
0.001	0.05 (1-tailed)	0.2	2	4	10728	42912
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	315268	315268
0.0001	0.05 (2-tailed)	0.2	2	1	235500	235500
0.0001	0.05 (2-tailed)	0.2	2	4	134194	536776
0.0001	0.05 (1-tailed)	0.2	2	4	107418	429672
0.0001	0.05 (2-tailed)	0.1	4	1	58376	58376
0.0001	0.05 (2-tailed)	0.2	4	1	43606	43606
0.0001	0.05 (2-tailed)	0.2	4	4	22572	90288
0.0001	0.05 (1-tailed)	0.2	4	4	18331	73324

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 α 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 Schlesselman):[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 above, p is the prevalence of the exposure in the control group, and K is the ratio of undiseased control subjects to diseased cases,

$$U = \frac{p}{K+1}K + \frac{R}{1+p(R-1)}$$

and

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

Again, a series of tables that provide sample sizes for case–control studies is presented in the Appendix. In Tables A9 through A12, we have assumed an α (two-tailed) of 0.05, a β of 0.1 (90% power), and control to case ratios of 1:1, 2:1, 3:1, and 4:1, respectively. Tables A13 through A16 are similar, except they assume a β 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 non-steroidal 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 α, β, 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 per diseased subject, assuming an α (two-tailed) of 0.05 and a β of 0.1? The sample size needed to detect a relative risk of 2.0 for any disease would vary, depending on the prevalence of use of the drug being studied. If one optimistically assumed 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 β 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 considers it inconceivable that this new drug could *protect* against liver disease, then one might use a one-tailed α, 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 α of 0.05, a β 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 the prior chapter, 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

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

Hypothetical drug prevalence rate assumed in undiseased	α	β	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	31588	31588
0.001	0.05 (2-tailed)	0.2	2	1	23596	23596
0.001	0.05 (2-tailed)	0.2	2	4	13449	53796
0.001	0.05 (1-tailed)	0.2	2	4	10765	43060
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	315373	315373
0.0001	0.05 (2-tailed)	0.2	2	1	235579	235579
0.0001	0.05 (2-tailed)	0.2	2	4	134240	536960
0.0001	0.05 (1-tailed)	0.2	2	4	107455	429820
0.0001	0.05 (2-tailed)	0.1	4	1	58409	58409
0.0001	0.05 (2-tailed)	0.2	4	1	43631	43631
0.0001	0.05 (2-tailed)	0.2	4	4	22585	90340
0.0001	0.05 (1-tailed)	0.2	4	4	18342	73368

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 10000 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 usually 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 a concern over whether it could cause agranulocytosis, since it was closely related

chemically to metiamide, 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 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 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 presents a table 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 non-steroidal 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 three. 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 the example above, it would be $0.206 \times 0.003 = 0.000618$. Analogously, the upper boundary would be the product of the corresponding "upper limit factor (U)" multiplied by the observed incidence rate. In the above example, this would be $2.92 \times 0.003 = 0.00876$. 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 no more than three 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 no more than three 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 one in 3333 subjects in the general population.

Discussion

The above discussions about sample size determinations in cohort and case–control studies assume one is able to obtain information on each of the five variables that factor into these sample size calculations. Is this in fact realistic? Four of the variables are, in fact, totally in the control of the investigator, subject to his or her specification: α, β, 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, this 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 α should be 0.05, as opposed to 0.06 or 0.04. The other variables

specified by the investigator 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 you 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 your population, in order to measure the information you need. 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 your sample size should be increased to take into account the imprecision of your 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 you expect 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 than expected 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 the investigator examines his or her final assumptions closely, asking whether, given the compromises made, the study is still worth undertaking.

Note that the discussion above 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 above, 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 automated 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: 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.

CHAPTER 5
When Should One Perform Pharmacoepidemiologic Studies?

Brian L. Strom
Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, 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 which one considers tolerable. Thus, this 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 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. Each will be discussed in turn.

Regulatory

Perhaps the most obvious and compelling reason to perform a postmarketing pharmacoepidemio-

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

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 re-positioning 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 whether:
      a. 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

logic 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 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] These are discussed more in Chapters 1 and 3. Now that the FDA has the authority to require such studies, such requirements are likely to become even more common.

Sometimes, a manufacturer may offer to perform a pharmacoepidemiologic study with the hope that the regulatory agency might thereby approve the drug's earlier marketing. 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, with the understanding that additional data would be gathered during postmarketing testing. For example, zidovudine was released for marketing after only limited testing, and only 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 that marketed a drug relatively early could be useful in demonstrating the safety of the drug to regulatory agencies in countries that have not yet permitted the marketing of the drug. 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 below, 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 clarifying any 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 re-position 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 38) and on patients' quality-of-life (see Chapter 39). 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 or in the elderly. By exploring unintended beneficial effects, or even drug efficacy (see Chapter 37), 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 avail-

ability 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 non-steroidal 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 they eventually were 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 post-marketing 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 can 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 above[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 above 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 37).

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 so-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, that is 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 32).

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 these drugs. Drugs used to treat mild or self-limited diseases also need careful study, 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 also 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 43.

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 versus just in sensitive people
   7. Known with certainty or not
B. Characteristics of the exposure
   1. Essential versus optional
   2. Present versus absent
   3. Alternatives available
   4. Risk assumed voluntarily
   5. Drug use will be as intended versus misuse is likely
C. Perceptions of the evaluator

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.

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 (see Chapter 3). Use of tampons is extraordinarily strongly linked to toxic shock: prior studies have shown relative risks 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 extraordinarily small risk of the 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 so-called "dread diseases," diseases that generate more fear and emotion than other diseases. 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, an 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, which 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 pharmacological 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 6-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 treatment of AIDS, despite its toxicity and the limited testing which 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 that 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 also 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 also 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 also was no question 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 generate 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[a]

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)	51.0
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

[a] Data derived from references 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 section of this 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 (accesses June 2011).
2. Humphries TJ, Myerson RM, Gifford LM, Aeugle ME, Josie ME, Wood SL, *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: 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, and 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, Spitz RW, Hennessy S, Feldman H, *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, Hennessy S, Kimmel SE, Farrar J, *et al*. Parenteral ketorolac: the risk for acute renal failure. *Ann Intern Med* 1997; **126**: 193–9.
12. Hennessy S, Kinman JL, Berlin JA, Feldman HI, Carson JL, Kimmel SE, *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. The 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

CHAPTER 6
The Role of Pharmacoepidemiology in the Health-Care System and Academia

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

Every year the health-care system has more medications at its disposal, each with its own efficacy, side effects, and cost. But when a new drug is introduced into practice, its benefit-to-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 fit into daily practice and into the health-care delivery system as a whole. 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, health-care systems all over the globe are struggling with how to provide the most appropriate care in the face of costs rising more quickly than inflation, and increasingly tight fiscal constraints. Pharmacoepidemiology can serve as a key tool for helping to address all of these concerns. These issues are growing in importance throughout the heath-care system, and are becoming particularly acute in academic medical centers.

Once a drug is approved for marketing, its prescription, its use by patients, and its outcomes traditionally move into a kind of "automatic pilot" status. Until recently, scant attention has been paid to systematic surveillance of these areas, except for the atypical settings of some integrated health-care delivery systems. The prevailing view has been that after the US Food and Drug Administration or comparable national authority approves a drug, it is used at the discretion of the physician, 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. It is only recently that more attention has been paid to assessing the outcomes of medication use on a population level, considering what its useful and harmful effects 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. This shift is being brought about by several factors. The first is the understanding that some adverse events can be identified and their risk adequately quantified only by observing its use in large numbers of patients. The second is the growing (though still nascent) sense that the best perspective on the impact of a medication on the health of the public requires measuring those outcomes in the health-care system itself, rather than one person at a time. It is here that the insights of pharmacoepidemiology can play an increasingly central role.

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

Driven by the pressures noted above, this situation has begun to change, 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 unnoticed or underappreciated at the time of approval; (ii) in typical practice, physicians often make prescribing decisions that do not reflect the best evidence-base or guideline recommendations; (iii) 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; (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.

The drug approval process

Each national health-care system must grapple with the following inherent paradox of pharmacology: A new therapy must be evaluated for approval at a time when the available data on its benefits and harms is 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. Traditionally, pharmacoepidemiology has played only a limited role at this stage, because of the widely accepted belief that only randomized controlled trials can provide the rigorous evidence of efficacy that is required for drug approval. However, since any medication that is effective is bound to have some adverse effect in some organ system in some patients, any approval must by definition be based on a judgment that a drug's efficacy is "worth it" in light of the known and unknown risks of the treatment. Yet the trials conducted by a given drug manufacturer to win approval are often powered (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 (e.g., a change in a laboratory test such as hemoglobin A1c or low density lipoprotein [LDL] cholesterol), the number of subjects required for these exercises is often inadequate to reveal important safety problems, if present. This is exacerbated by the extensive exclusion criteria for study participation, and the often brief duration of these trials.[1]

As a result, additional methods often need to be applied even to preapproval data to aggregate adverse events from multiple study populations to provide the power needed to assess safety. If one extends the definition of pharmacoepidemiology to include such systematic aggregation and meta-analysis (see Chapter 40) of adverse effects data from multiple preapproval trials, this represents the first opportunity to use these tools to inform the appropriate use of medications. An example of this is the combining of 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),[2] rosiglitazone (myocardial infarction),[3] or the selective serotonin reuptake inhibitors (SSRIs) used in children (suicidality).[4] Thus, the methods of pharmacoepidemiology can begin to play a role in addressing drug safety questions to inform policy and utilization decisions early in a drug's life cycle.

Prescribing practices

Once a drug has entered the health-care delivery system, a growing literature documents several areas in which prescribing often falls short of existing knowledge. These issues as well can 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 adequately 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.[5] Even utilizing a database that includes only drug utilization information, pharmacoepidemiology makes it possible to achieve a good first approximation of the problem of under-treatment by measuring the age- and gender-adjusted prevalence of use of medications to manage specific chronic conditions by a given physician, or a given practice or health system (see Chapter 24). When patterns of use are combined with other research on physician characteristics and decision making, it becomes possible to identify more clearly when and how prescribing is likely to fall short, insights which can then be used to shape programs to improve care (see below).[6]

When medications are used, there is good evidence that physicians frequently do not prescribe regimens that are optimal, based on the available clinical evidence, or they 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 doctor, 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.

If diagnostic data are also available, as they increasingly are in many health-care systems, a more sophisticated approach can also take into account contraindications and compelling indications related to specific drug choices, in order to refine the assessment of the appropriateness of prescribing in an entire health-care system, or for individual clinicians (see Chapter 25). Numerous studies have documented 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 on Prevention, Detection, and Treatment of High Blood Pressure (JNC).[7] This study found that a substantial proportion of treated hypertensive patients were not receiving a regimen consistent with JNC guidelines. Often, such suboptimal prescribing involved omission 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.[8]

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 or data from the visit itself to measure the adequacy of use of cholesterol-lowering agents, diabetes drugs, or antihypertensives. This makes it possible to assess the effectiveness of prescribing outcomes for a given physician (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 health-care system

Beyond the problem of under- or mis-prescribing by physicians, under-use of needed drugs by patients is one of the most common medication-related

problems, and one that can be readily identified by pharmacoepidemiology.[9] Although it is less striking than obvious drug-induced adverse events, under-use 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 health-care systems.[10] Thus, the clinical trialist or the treating physician focuses respectively on patients who are assigned to receive a drug in a study, or who are prescribed a drug in practice. 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 the same diagnoses 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 (see Chapter 42) 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 "non-compliant" 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 regimen over time. *Intelligent non-adherence* (or intelligent non-compliance) describes a situation in which a patient stops a therapy because it is producing excessively burdensome side effects or failing to relieve symptoms effectively.

The field of modern adherence research (see Chapter 42) is relatively new, because assessing patient compliance with prescribed medications on a large scale required the advent of computerized pharmacy claims datasets to make such measurements efficiently. Until around 1990, this was an under-studied area, and most physicians assumed that after they wrote a prescription, a patient filled it and took it more or less as directed. But once 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.[11,12]

Datasets based on the complete paid-claims files of drug benefit programs provided the first means of studying adherence in defined populations. Because such a claim is necessary before the pharmacy can be paid, and because insured patients may be unlikely to pay out-of-pocket to fill prescriptions outside the system, such datasets provided an excellent record of what medications are actually dispensed.[13] 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, etc. are not taken.[14] This causes a massive and still under-appreciated 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.[15] The full magnitude of this problem is still not completely appreciated by clinicians or policymakers.

Because many assessments of under-use are based on pharmacy-generated data on filled prescriptions, it is sometimes difficult to know whether non-use of an indicated drug was the result of a failure of the patient to fill a prescription, or the failure of the physician to write it. The advent of electronic prescribing is making 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.[16] One large study found that a fourth of initial prescriptions written electronically were never picked up at the pharmacy.[17] As a result, the approximately 50% rate of non-adherence seen over time in the 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 even initiated by the patient. Terminology has evolved to define the two aspects of this problem: *secondary non-adherence* refers to the failure by a patient to continue on a drug regimen that he or she has already begun; *primary non-adherence* occurs when a physician writes a pre-

scription that the patient does not even fill for the first time.

These findings about adherence have implications for other aspects of pharmacoepidemiologic studies, described in other chapters. First, they raise important concerns about the validity of large databases (such as the UK General Practice Research Database) that define drug exposure in terms of what a doctor prescribed, as opposed to what the patient actually obtained from the pharmacy (see Chapter 15). Second, the very high rates of non-use 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 32 and 37).

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 non-adherent.[18] 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.[19] Another consistent risk factor has been race, suggesting an important problem in physician–patient communication and/or trust for non-white patients.[20] But 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

An application of the tools of pharmacoepidemiology that is certain to see more widespread use in the coming years is the assessment of the outcomes of medication use in typical "real-world" populations. This perspective is based on the difference between *efficacy*, which is the effect of a medication in the rigorous but idealized setting of a clinical trial, compared to its *effectiveness*, which is a measure of its outcomes in typical practice settings (see Chapters 32 and 37). These often differ. For example, one important conventional randomized trial demonstrated convincingly that addition of spironolactone to the regimen of patients with congestive heart failure substantially improved their clinical status and reduced mortality.[21] However, a population-based time-series analysis later found 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.[22] By contrast, an analysis of prescribers' response to a different study that provided new evidence about optimal management of atrial fibrillation demonstrated a more positive change in practice.[23]

Other "lost in translation" problems 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.[24] Such under-use was found to be associated with physicians' recent experience with adverse events caused by the drug,[25] as well as by their perceptions and attitudes about risks and benefits.[26] 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.[27] This kind of real-world population research can lay the foundation for enlightened interventions to address such non-use, by addressing its underlying causes.

Pharmacoepidemiologic methods can also be used to track the diffusion of new medication classes into practice,[28] as well as the reaction of practitioners in various settings to new information

about drug risks, as in the case of warnings about the cardiovascular toxicity of rosiglitazone.[29]

Policy analysis

Usually, policy changes are implemented in the health-care 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 health-care 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 health-care budget.[30] Such observational methods have also been combined with population-based randomized policy trials, and were found to yield similar results.[30]

Similarly, one large U.S. 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.[31]

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.[32,33] The methods of pharmacoepidemiology are increasingly used to assess the clinical and economic consequences of such policies.[34–36] One study documented an increase in use of clopidogrel in one Canadian province after a highly restrictive policy was replaced with a more lenient one; importantly, this change was associated with a significant concomitant reduction in adverse cardiovascular outcomes.[37]

Interventional pharmacoepidemiology

Once the tools of pharmacoepidemiology make it possible to define patterns of suboptimal use, such as poor drug choices, under-use, over-use, and problematic dosing, such surveillance can be employed to identify problems that may be 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 25. This approach was designed to build on 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 this 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 proactively 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.[38] 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.[39–41]

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 programs' pharmacy datasets made possible a rigorous assessment of their 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.[42] In the US, the largest of these is the non-profit Independent Drug Information Service (www.RxFacts.org), which serves state-funded health-care programs in several states.

Economic assessment of medication-related issues

Once assessments of drug use are performed in a population-based dataset that contains information on expenditures as well as utilization, it is possible to assess the economic impact of such prescribing issues as well (see also Chapter 38). The above study of patients treated for hypertension, for example, found that better adherence to the guideline recommendations of JNC 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.[7] 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.[8]

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.[43] For example, although statins usually must be taken for a lifetime, the randomized trials demonstrating their benefit often last for a much shorter time, often under 2 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 that should be both transparent and conservative—it becomes possible to estimate the lifelong benefits, risks, and costs of use of such treatments.[44]

The academic medical center

The academic medical center represents a special case of inquiry for pharmacoepidemiology, and one where it 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. Although the patients cared for in the acute hospital setting comprise a temporary population, data on the medications they receive and the outcomes that result can nonetheless benefit from the perspectives and tools of pharmacoepidemiology. For example, one study assessed the use of erythrocyte stimulating agents such as erythropoietin to define how it was being used, as well as the adequacy of concurrent treatment with other required therapies, such as iron.[45] Similarly, academic medical centers can provide an opportunity to study medication-related adverse outcomes in patients undergoing invasive procedures. For example, conventional pharmacoepidemiologic methods make it possible to define cohorts of cardiac surgery patients exposed to agents such as aprotinin or hetastarch, and then observe peri- and postoperative outcomes including hemorrhage and other postoperative morbidity and mortality. In each of these cases, pharmacoepidemiology applied to routine use of inpatient medications made it possible to define serious adverse events whose frequency and severity had been previously under-appreciated.[46,47]

The application of population-based approaches can then 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 45).[48] Until recently, this was possible only in advanced comprehensive health-care organizations. But 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.[49] 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.[40] 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 health-care 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. 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.[50] This consortium has several features that hold promise of advancing the use of pharmacoepidemiology in the health-care setting. 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[51] 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.[52] It is anticipated that ENCePP will also make study designs and data available publicly, to enable others to scrutinize methods and even re-analyze study data. This should help to increase the standards for pharmacoepidemiologic research globally.

In the United States, the federal Agency for HealthCare Research and Quality (AHRQ) has established the DEcIDE network (Developing Evidence to Inform Decisions about Effectiveness), to help inform policymakers about the comparative

effectiveness and safety of alternative treatment approaches (see also Chapter 32 on comparative effectiveness research). It has established a particular focus on methodologic issues, since they play such a critical role in the validity and generalizability of non-randomized comparative effectiveness studies. The CERTs program (Centers for Education and Research in Therapeutics) is another AHRQ program, designed to support academic programs seeking to study medication risks and benefits, and to optimize the use of those medications.

Other examples in the United States of consortia of delivery systems and/or academic centers to further pharmacoepidemiologic research are the Sentinel Initiative of the FDA, mandated to perform postmarketing surveillance of adverse events (see Chapter 30), and the HMO Research Network (see Chapter 12). These are described in later chapters.

A major milestone in the real-world application of pharmacoepidemiologic methods in the US health-care system was the establishment in 2010 of the Patient-Centered Outcomes Research Institute (PCORI). A product of the health-care reform program enacted the same year, PCORI was designed to be a stable source of ongoing funding for comparative effectiveness research, which will largely involve the study of medications, often by means of observational studies (see also Chapter 32).[53]

The future

The continuing evolution of health-care systems in both the industrialized and the developing worlds will likely bring about a growing role for pharmacoepidemiology in a wide variety of settings. Continuing research on the fundamental mechanisms of disease is leading to new medications of unprecedented efficacy that also carry daunting risks of toxicity, and often enormous costs. Health-care systems all over the world face pressures as never before to provide only those interventions that have the best efficacy and safety, but also at the most affordable price. To accomplish this will require a reliance 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 finding the most inexpensive medication necessarily lead to optimal outcomes clinically, economically, or ethically. Pharmacoepidemiology (and its spinoff discipline, pharmacoeconomics) bring to an increasingly challenged health-care system the tools that make possible rigorous quantitative 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.

The development of new approaches and programs in this emerging discipline, and their growing uptake to guide real-world decision making at both the clinical and policy levels, bode well for the evolution of health-care delivery into a more data-driven enterprise. If these encouraging trends continue, it will mark the transition of pharmacoepidemiology from a small, under-populated discipline at the fringe of the worlds of health-care delivery and academic medicine, to a vibrant, growing discipline at their very center.

References

1. Avorn J. *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*. New York: Alfred A. Knopf, 2004.
2. 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**: 1976–85.
3. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007; **356**: 2457–71.
4. Williams SB, O'Connor E, Eder M, Whitlock E. Screening for child and adolescent depression. In: *Primary Care Settings: A Systematic Evidence Review for the U.S. Preventive Services Task Force* [Internet]. Rockville, MD: Agency for Healthcare Research and Quality (US), 2009.
5. 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**: 1061–7.

6. 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.
7. Fischer MA, Avorn J. Economic implications of evidence-based prescribing for hypertension: can better care cost less? *JAMA* 2004; **291**: 1850–6.
8. Choudhry NK, Levin R, Avorn J. The economic consequences of non-evidence-based clopidogrel use. *Am Heart J* 2008; **155**: 904–9.
9. Cutrona SL, Choudhry NK, Stedman M, Servi A, Liberman JN, *et al*. Physician effectiveness in interventions to improve cardiovascular medication adherence: a systematic review. *J Gen Intern Med* 2010; **25**: 1090–6.
10. 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**: 548–54.
11. Gurwitz JH, Glynn RJ, Monane M, Everitt DE, Gilden D, Smith N, *et al*. Treatment for glaucoma: adherence by the elderly. *Am J Public Health* 1993; **83**: 711–6.
12. 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**: 697–704.
13. Avorn J, Monette J, Lacour A, Bohn RL, Monane M, Mogun H, *et al*. Persistence of use of lipid-lowering medications: a cross-national study. *JAMA* 1998; **279**: 1458–62.
14. Jackevicius CA, Mamdani M, Tu JV. Adherence with statin therapy in elderly patients with and without acute coronary syndromes. *JAMA* 2002; **288**: 462–7.
15. Jackevicius CA, Li P, Tu JV. Prevalence, predictors, and outcomes of primary nonadherence after acute myocardial infarction. *Circulation* 2008; **117**: 1028–36.
16. Hakim A. E-prescribing and primary noncompliance: physician and patient experience. *Prof Case Manag* 2010; **15**: 62–7.
17. Fischer MA, Stedman MR, Lii J, Vogeli C, Shrank WH, Brookhart MA, *et al*. Primary medication nonadherence: analysis of 195,930 electronic prescriptions. *J Gen Intern Med* 2010; **25**: 284–90.
18. Chan DC, Shrank WH, Cutler D, Jan S, Fischer MA, Liu J, *et al*. Patient, physician, and payment predictors of statin adherence. *Med Care* 2010; **48**: 196–202.
19. Shrank WH, Hoang T, Ettner SL, Glassman PA, Nair K, DeLapp D, *et al*. The implications of choice: prescribing generic or preferred pharmaceuticals improves medication adherence for chronic conditions. *Arch Intern Med* 2006; **166**: 332–7.
20. Hill TD, Graham LM, Divgi V. Racial disparities in pediatric asthma: a review of the literature. *Curr Allergy Asthma Rep* 2011; **11**: 85–90.
21. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, *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**: 709–17.
22. Juurlink DN, Mamdani MM, Lee DS, Kopp A, Austin PC, Laupacis A, *et al*. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med* 2004; **351**: 543–51.
23. Choudhry NK, Zagorski B, Avorn J, Levin R, Sykora K, Laupacis A, *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**: 1563–72.
24. 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**: 978–84.
25. 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**: 141–5.
26. 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**: 1060–5.
27. 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**: 1111–5.
28. 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**: 1819–28.
29. 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.
30. 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**: 560.

31. Choudhry NK, Fischer MA, Avorn J, Schneeweiss S, Solomon DH, Berman C, *et al*. At Pitney Bowes, value-based insurance design cut copayments and increased drug adherence. *Health Aff (Millwood)* 2010; **29**: 1995–2001.
32. 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**: 2187–94.
33. Polinski JM, Wang PS, Fischer MA. Medicaid's prior authorization program and access to atypical antipsychotic medications. *Health Aff (Millwood)* 2007; **26**: 750–60.
34. 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.
35. Fischer MA, Servi AD, Polinski JM, Wang PS. Restrictions on antidepressant medications for children: a review of Medicaid policy. *Psychiatr Serv* 2007; **58**: 135–8.
36. 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**: 4–9.
37. Jackevicius CA, Tu JV, Demers V, Melo M, Cox J, Rinfret S, *et al*. Cardiovascular outcomes after a change in prescription policy for clopidogrel. *N Engl J Med* 2008; **359**: 1802–10.
38. 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.
39. Avorn J, Soumerai SB, Everitt DE, Ross-Degnan D, Beers MH, Sherman D, *et al*. A randomized trial of a program to reduce the use of psychoactive drugs in nursing homes. *N Engl J Med* 1992; **327**: 168–73.
40. Solomon DH, Van Houten L, Glynn RJ, Baden L, Curtis K, Schrager H, *et al*. Academic detailing to improve use of broad-spectrum antibiotics at an academic medical center. *Arch Intern Med* 2001; **161**: 1897–902.
41. Soumerai SB, Avorn J. Economic and policy analysis of university-based drug "detailing". *Med Care* 1986; **24**: 313–31.
42. O'Brien MA, Rogers S, Jamtvedt G, Oxman AD, Odgaard-Jensen J, Kristoffersen DT, *et al*. Educational outreach visits: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2007; **4**: CD000409.
43. 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**: 780–7.
44. 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 Amer Coll Cardiol* 2011; **57**: 784–91.
45. 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.
46. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med* 2008; **358**: 771–83.
47. Avorn J, Patel M, Levin R, Winkelmayer WC. Hetastarch and bleeding complications after coronary artery surgery. *Chest* 2003; **124**: 1437–42.
48. Schedlbauer A, Prasad V, Mulvaney C, Phansalkar S, Stanton W, Bates DW, *et al*. What evidence supports the use of computerized alerts and prompts to improve clinicians' prescribing behavior? *J Am Med Inform Assoc* 2009; **16**: 531–8.
49. 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**: 647–53.
50. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. http://www.encepp.eu/ (accessed June 2011).
51. European Network of Centres for Pharmacoepide miology and Pharmacovigilance. *Code of Conduct*. http://www.encepp.eu/code_of_conduct/index.html (accessed June 2011).
52. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. *Standards and Guidances*. http://www.encepp.eu/standards_and_guidances/index.html (accessed June 2011).
53. Clancy C, Collins FS. Patient-Centered Outcomes Research Institute: the intersection of science and health care. *Sci Transl Med* 2010; **2**: 37cm18.

CHAPTER 7
The Role of Pharmacoepidemiology in Industry

Jingping Mo, Nicolle M. Gatto, Rachel E. Sobel, and Robert F. Reynolds
Pfizer Inc., New York, NY, USA

Introduction

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 product planning, portfolio development, and the commercialization of drugs. The use of epidemiology to support the commercialization and appropriate marketing of drugs, including studies of beneficial drug effects, health economics, and quality-of-life measures, are discussed elsewhere in this book (see Chapters 37 to 39). The most visible contribution of epidemiology in the biopharmaceutical industry is arguably drug safety evaluation, including the contextualization of safety signals and examination of specific research hypotheses. To meet these aims, epidemiologists design and implement background epidemiologic studies among indicated populations, risk management interventions and evaluations, and postapproval safety studies. Additionally, epidemiologists contribute strategy, content, and expertise to 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 lifecycle from the perspective of epidemiologists working within the biopharmaceutical industry.

Regulatory and industry focus on risk management and epidemiology

Biopharmaceutical risk management is fundamentally concerned with preserving a favorable benefit–risk balance in 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 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 views expressed are those of the authors, which are not necessarily those of Pfizer Inc.

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

The evolution of biopharmaceutical risk management

Between food, drugs, and cosmetics, it is estimated that the US FDA is responsible for regulating 25% of the US gross national product.[1,2] Safety has been a central theme in biopharmaceutical regulation since the first Food and Drug Act of 1906, which prohibited the manufacture and sale of mislabeled or adulterated drugs. It was not until 1938 however that manufacturers and marketers were required to demonstrate the safety of drugs as a result of more than 100 deaths from an adulterated sulfanilamide preparation. Amendments in 1962 extended the requirement to both efficacy and safety after birth defects were found to be associated with the use of thalidomide. In the intervening years, the FDA handled risk management type activities on a case-by-case basis, such as requiring manufacturers to communicate to prescribers and patients via Dear Healthcare Professional Letters or Patient Package Inserts, respectively. The first restricted distribution product came in 1990 with clozapine, in which patients could not receive a prescription until safe-use conditions, i.e., no agranulocytosis, were demonstrated (the "no blood no drug" campaign).

Public pressure to speed drug approvals for HIV and cancer drugs led to the Prescription Drug User Fee Act (PDUFA). Ten years later, the 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 of 2002. For the first time, dedicated funding was provided to the FDA for risk management resources. In response to this regulation, the FDA issued three guidances in 2005 on: (i) Premarketing Risk Assessment, (ii) Pharmacovigilance and Pharmacoepidemiology, and (iii) Risk Minimization Action Plans (RiskMAPs). By 2007, at least 16 products had RiskMAPs, which could include any combination of enhanced education, reminder systems, or performance-linked access, such as restricted distribution, to minimize risks and maintain a positive benefit–risk profile in appropriate patient populations.

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 to assess and approve 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 focused on the US FDA's structure and function, but also assessed the role of the biopharmaceutical industry, academia, the health-care system, the legislature, patients, and the public. The IOM committee made numerous recommendations, several of which pertained to epidemiologists. The IOM recommended that the FDA receive additional funding and staff; improve communications on drug safety, including a larger role for the drug safety staff; and, most importantly, be given additional authority and enforcement tools.[3]

As a result of the IOM report and other stakeholder research and advocacy, Congress passed the Food and Drug Administration Amendment Act FDAAA 2007 (PL 110-85). The FDAAA further strengthened the FDA's oversight of risk management activities (see Chapter 29). Previously, RiskMAPs and postapproval commitment studies were defined agreements between industry and the Agency. Although the FDA frequently required such studies of sponsors of new drugs, its legal authority to require and enforce completion of these activities postapproval was perceived by many to be limited. With FDAAA, the FDA was granted the ability to mandate postapproval studies (postmarketing requirements, or PMR) and risk mitigation evaluation strategies (REMS; see section later in this chapter for further information) by imposing substantial fines for non-compliance 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 guidance issued in 2011.[4]

Europe had similar legislation passed 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). EU law requires companies to submit a formal risk management plan (RMP) with each marketing authorization application (MAA). The EU defines risk management as a set of pharmacovigilance activities and interventions designed to identify, characterize, prevent, or minimize risks relating to medicinal products, including the assessment of the effectiveness of those interventions. The RMP contains two key sections: (I) Safety Specification/Pharmacovigilance Plan and (II) Evaluation of the Need for Risk Minimization Activities, and the associated plan if applicable. Safety Specifications include known and hypothetical risks, or areas of missing information. The RMP also requires the MAA to provide extensive background epidemiologic data on the incidence and prevalence of the risks (safety specifications) in the underlying disease population, in addition to the incidence and prevalence of the disease itself. The Pharmacovigilance Plan section describes any planned and ongoing studies, for example pharmacoepidemiologic or non-clinical studies, or enhanced surveillance activities. In addition, Volume 9A defines the criteria and reporting obligations for "post authorization safety studies," or PASS, which includes most postapproval epidemiologic studies to monitor the safety of drugs in actual clinical practice. Finally, the "Evaluation" section of the RMP summarizes any risk minimization activities and their evaluation plans, similar to REMS in the US. While epidemiology has become increasingly important to risk management over the last three decades, FDAAA and Volume 9A have further solidified epidemiology's role in informing the benefit–risk 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 the postapproval lifecycle 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 is 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. 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 provide a more comprehensive and pragmatic picture of the safety profile of a drug as it is used in clinical practice.

Pharmacoepidemiologic study designs

Pharmacoepidemiologic studies can be descriptive or analytic in nature, may involve existing data or primary data collection, and may be used to generate 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), and *de novo* observational studies (see Part III of this book). Commonly used study designs include the prospective or retrospective cohort study, case–control study, or cross-sectional study (see Chapter 3). To address any specific product safety concern, it is important to consider all potential study design options before choosing the most appropriate one and implement epidemiologic study in accordance with the *Guidelines for Good Pharmacoepidemiology Practices* (GPP).[5]

In addition to typical epidemiologic designs, depending on the specific safety research hypothesis, epidemiologists design and implement enhanced surveillance studies, large simple trials, and case-crossover studies. Enhanced 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 enhanced surveillance program established for juvenile idiopathic arthritis can be found later in this chapter in the section on pediatrics.

A large simple trial (LST) is a hybrid design that combines randomization to treatment with observational follow-up of patients (see Chapter 36). This design allows for 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, the LSTs approximate real life practice. Further, the large study size provides the power needed to evaluate small absolute and relative risks. An example of a LST is illustrated in ziprasidone's ZODIAC Large Simple Trial in Case study 7.1.

The case-crossover study design, which is analogous to a traditional matched case–control design and was developed to assess effects of intermittent exposures on diseases with abrupt onset, is currently being used to evaluate whether PDE5 inhibitors (i.e., sildenafil, vardenafil, and tadalafil), as a class, trigger the onset of acute non-arteritic anterior ischemic optic neuropathy (NAION).[14] It would be very challenging to identify appropriate controls with a standard case–control study design. Thus, the case-crossover study was considered the preferred design as compared to the matched case–control study since: PDE5 inhibitors are taken on an as-needed basis, which constitute 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 himself, so that the potential effects of confounders that vary over long periods of time, such as age, diabetes, and hypertension, are effectively held constant.

Data sources for pharmacoepidemiologic 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, has experienced considerable growth over the past two decades (see Part IIIB of this book). These databases offer several advantages over *de novo* epidemiologic studies or expanding the scope of clinical trials. 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 results of studies in which patients are identified and followed over time. The third advantage is cost relative to prospective studies. Clinical trials or other prospective observational studies may cost 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 and its coding systems; the resources devoted to evaluating and monitoring the research quality of the data; and national or regional variations in medical practice (see also Chapter 11). A common research limitation of automated data sources is that sufficient number 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 have overall enrollments of hundreds of thousands of members, these numbers may be inadequate to study the risks of extremely rare events associated with a specific drug or the drug may not be contained in the HMO's limited research database. Finally, results from these sources are often limited in their generalizability.

Case study 7.1 Risk management: ziprasidone

The FDA approved ziprasidone for the treatment of schizophrenia in February 2001. Initially, in June 1998, the Sponsor (Pfizer Inc.) of the ziprasidone NDA received a not approvable letter based on "the judgment that ziprasidone prolongs the QTc and that this represents a risk of potentially fatal ventricular arrhythmias that is not outweighed by a demonstrated and sufficient advantage of ziprasidone over already marketed antipsychotic drug products".[6] The letter of non-approval recommended that the Sponsor perform an additional study to determine the QTc effect of ziprasidone at peak plasma concentration in comparison with other atypical antipsychotics and with several standard antipsychotics.[6]

The Sponsor conducted a comparative clinical study of six antipsychotics which indicated ziprasidone's QT_c interval at steady state was 10 milliseconds greater than that of haloperidol, quetiapine, olanzapine, and risperidone and approximately 10 milliseconds less than that of thioridazine; further, the results were similar in the presence of a metabolic inhibitor.[7] Following the 1998 non-approval, the sponsor also conducted descriptive and comparative epidemiologic studies to quantify the risk of mortality and cardiovascular disease among schizophrenic patients receiving pharmacotherapy, and conducted an innovative postapproval study for assessing the safety of ziprasidone.

Descriptive epidemiologic studies

Numerous studies have documented that patients with schizophrenia have higher mortality rates than the general population but few had examined if these rates changed following the introduction of atypical antipsychotics. Prior to approval, and as part of the ziprasidone epidemiology program during the drug's development, two descriptive epidemiologic studies were conducted: one in the US used United Healthcare's Research Database[8] and another in Canada used Saskatchewan Health's database.[9] The results confirmed that patients with schizophrenia have higher background rates of mortality and cardiovascular outcomes, regardless of treatment type.

Comparative epidemiologic studies

In an effort to determine the "real-world" effects of the use of QT_c prolonging drugs among patients with schizophrenia, the sponsor also conducted two comparative epidemiologic studies of antipsychotics already on the market using data from the US Medicaid system[10] and the UK General Practice Research Database.[11] These studies compared antipsychotics with varying propensities for QT_c prolongation, from lower to higher: haloperidol, risperidone, clozapine, and thioridazine. The results indicated that rates of sudden death and cardiac events are similar among users of haloperidol, clozapine, risperidone, and low-dose thioridazine, but that users of high-dose thioridazine have higher rates of these events.

Postapproval safety study: ZODIAC large simple trial

The Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC) was a large simple trial designed to examine the "real-world" cardiovascular safety of ziprasidone compared with olanzapine.[12] The defining characteristics of ZODIAC included:

- Prospective study large enough to detect small risks: 18154 patients from 18 countries: Argentina, Brazil, Chile, Hong Kong, Hungary, Korea, Malaysia, Mexico, Peru, Poland, Romania, Singapore, Slovakia, Sweden, Taiwan, Thailand, Uruguay, and the United States.
- Control for confounding by indication with 1:1 random assignment to ziprasidone or olanzapine.
- No additional study monitoring or tests required after randomization.
- Patients followed up during usual care over 12 months regardless of how long the patient stayed on randomized medication.
- Primary endpoint: non-suicide mortality. Secondary endpoints: all-cause mortality, mortality due to suicide, cardiovascular mortality, mortality due to sudden death, all-cause hospitalization, hospitalization for arrhythmia, myocardial infarction, or diabetic ketoacidosis, and rate of discontinuation of randomized treatment.
- Three independent scientific committees: a Scientific Steering Committee responsible for general oversight of the study, a Data Safety Monitoring Board safeguarding study participants, and an Endpoint Committee, blinded to treatment status, and charged with assessing whether reported events meet study endpoint criteria.

The primary analyses found no difference between the ziprasidone and olanzapine treatment arms with respect to non-suicide mortality (RR = 1.02, 95% CI: 0.79–1.39). 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.[13]

The design of ZODIAC carries several advantages over more commonly used observational postmarketing study designs. Random allocation of patients provides for an unbiased comparison between groups; the large study size provides the power needed to evaluate small risks, both absolute and relative; and the simplicity of an uncontrolled trial minimizes the artificiality imposed by controlled premarketing trials.

Key points:

- Epidemiologic studies can be used to establish baseline rates of disease in the patient population and compare the rates of adverse effects of drugs used for the same indication.
- Prospective epidemiologic studies, such as large simple trials, can be used to evaluate potentially small risks in a "real-world" context.

Many of these data collection systems were designed for administrative purposes, rather than for epidemiologic research studies. As a result, information needed to assess a specific safety issue may be unavailable and the quality of medical information may be inadequate. Often it is desirable to validate findings based solely on diagnostic or procedural codes used for reimbursement purposes through a detailed review of at least a subset of medical records, 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 because of concerns about patient confidentiality or anonymity, especially following enacted legislation on the privacy of health records (e.g., the Health Information Portability and Accountability Act, HIPAA). Continuing studies of the research validity of these databases is crucial, and should be pursued when feasible.[15–19] Further information on specific data sources can be found in Chapters 11–18 of this book.

Epidemiologic studies with primary data collection are considered when it is not feasible to address safety issues using existing databases. The NAION case-crossover study described earlier, the International Men's Health Study conducted for sildenafil (see below), the pegaptanib prospective cohort study in Europe (Case study 7.2), and the ziprasidone large simple trial (Case study 7.1) illustrated in this chapter are examples of epidemiologic studies that involved primary data collection. Obviously, these types of studies take a relatively long time for data collection and are more costly. In some instances, it is possible to identify patients or physicians in automated databases and then collect additional information directly from patients or physicians through phone interviews or mailed questionnaires to supplement data from existing data sources.[20,21] This is an efficient approach by combining secondary and primary data collection.

Contributions of preapproval epidemiology

Before evaluation of a potential medicine can begin, extensive preclinical research is conducted, 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 provide abundant data about identified and hypothetical risks, but still have several limitations. Preapproval randomized clinical trials 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 sufficiently large to provide evidence of a beneficial clinical effect, to exclude large increases in risk of common adverse events, and to 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. Typically,

Case study 7.2 Risk management: pegaptanib

Pegaptanib was the first ocular antivascular endothelial growth factor (VEGF) therapy approved in the US in 2004 and in the EU in 2006 for the treatment of all forms of neovascular age-related macular degenerations (AMD), the leading cause of vision loss in people 55 years of age and older in the developed world.[22] The drug product is a sterile, aqueous solution supplied in a single-dose, prefilled syringe, administered every 6 weeks as an intravitreal injection into the eye. Endophthalmitis is an infrequent but known serious adverse event associated with intravitreal injection procedure, rather than pegaptanib.[23] The reported incidence of endophthalmitis in the pegaptanib clinical development program is comparable with the rates observed with other intraocular injection drugs.[23]

Before marketing approval, the Sponsor (Pfizer Inc.) and regulators suspected that the incidence of endophthalmitis might be higher when pegaptanib was used in clinical practice since detailed instructions on how to perform intravitreal injection procedure were provided to investigators in the pegaptanib clinical development program, whereas only the approved pegaptanib label would be available to ophthalmologist as reference when the drug is used in the real world. As part of the pegaptanib epidemiology program conducted during its development, the Sponsor initiated one epidemiologic study and designed two epidemiology studies to quantify the risk of endophthalmitis among AMD patients who received intravitreal injections both before and after pegaptanib approval.

Pre- and postapproval epidemiologic studies in the US

Two complementary epidemiologic studies, designed to evaluate the trend of endophthalmitis incidence associated with intravitreal injection among AMD patients before (2000–2003) and after (2005–2006) pegaptanib became commercially available were conducted using Medicare databases in the US. Medicare data were considered the most appropriate existing data source as it is representative of the US population aged 65 years and older with approximately 98% of elderly enrolled in Medicare. Since AMD is predominantly a disease of the elderly, the study population is representative of AMD patients treated with intravitreal injection in the US.

It was found that the frequency of intravitreal injections for the treatment of AMD increased significantly from 2000 to 2006 (973 injections in 2000 and 2001, 4678 injections in 2002, 14 056 injections in 2003, 206 616 injections in 2005, and 494 514 injections in 2006) and the incidence of intravitreal injection-related endophthalmitis decreased significantly over the same period (1.92, 1.42, 0.42, 0.16, and 0.12 per 100 injections in 2000–2001, 2002, 2003, 2005, and 2006).[24,25] These findings suggested that with the introduction of intravitreal injection of anti-VEGF for the treatment of neovascular AMD, the incidence of endophthalmitis is low. The study investigators indicated that the most likely explanation for the decreased incidence is that ophthalmologists became more skilled in performing the intravitreal injection procedure as intravitreal anti-VEGF agents have become the standard of care for the treatment of neovascular AMD and more practicing ophthalmologists used standardized injection procedure following the introduction of intravitreal injection guidelines.[26]

Postapproval epidemiologic study in Europe

A similar, but *de novo*, epidemiology study was conducted in Europe to quantify the risk of endophthalmitis among pegaptanib-treated patients at the time of approval in the EU in 2006. This study was necessary since ocular medications were not routinely administered in Europe via an intravitreal injection at that time and there were differences in performing intravitreal injection procedures between Europe and the US. However, the Sponsor was unable to identify an existing database with ophthalmology specialty care data in Europe. Therefore, the Sponsor designed and implemented a prospective epidemiologic cohort study with primary data collection to estimate the incidence of endophthalmitis and other ocular events in patients receiving pegaptanib injections for neovascular AMD in 13 European countries, including Belgium, Cyprus, Czech Republic, Denmark, France, Germany, Greece, Ireland, Italy, Poland, Slovakia, Spain, and Sweden. Approximately 500 patients treated with pegaptanib for AMD at ophthalmic hospital clinics, private ophthalmic practices, and academic ophthalmic centers have been enrolled and these patients are now being followed for up to 2 years.

Key points:

- Early planning for and implementation of epidemiology studies during the development program provided estimates of the risk of adverse events before and after the commercially available of a new treatment.
- Epidemiology studies with primary data collection should be considered when existing data sources are not available to estimate the risk of event of interest in the indicated patient population.

these trials have a total patient sample size up to several thousands. 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.) While clinical trials are not intended or designed to address all potential safety issues related to a particular drug,[27] 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; expected prevalence/incidence of risk factors, co-morbidities and complications; patterns of health-care utilization and prescribing of currently approved treatments; and background rates of mortality and serious non-fatal events. Epidemiologists use this information to complete epidemiologic sections of key regulatory documents such as risk management and pediatric investigation plans and orphan drug applications.

Epidemiologic studies conducted before or during the clinical development program are also useful 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.[28–30] These background epidemiologic data can 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 next phase of development, risk management and mitigation planning, and regulatory approvals.

As an example, two preapproval descriptive studies were conducted to support the filing of ziprasidone for treatment of schizophrenia (see Case study 7.1). While studies had previously shown that patients with schizophrenia have higher rates of mortality than the general population, it was unknown whether these rates changed following the introduction of atypical antipsychotics. These descriptive database studies confirmed that patients with schizophrenia have higher background rates of mortality and cardiovascular outcomes, and provided more appropriate background mortality rates for patients likely to take ziprasidone.

To support pegaptanib filling for the treatment of all forms of neovascular age-related macular degeneration (AMD), two complementary studies were designed and proposed to regulators prior to approval. These studies were conducted in Medicare databases (see Chapter 14) to evaluate the trend of the endophthalmitis incidence associated with intravitreal injection among AMD patients before and after pegaptanib became commercially available in the US (see Case study 7.2).

In addition to summarizing the existing relevant literature and designing and executing background epidemiologic 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 epidemiologic studies during development. Planning for successful postapproval epidemiologic studies often begins well before approval. During the preapproval phase, epidemiologists may conduct feasibility assessments for planned postapproval studies, start 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).

Contributions of postapproval epidemiology

The need for a postapproval epidemiologic 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, spontaneous reports, published case series, or signal detection of electronic health-care data. Postapproval, epidemiologists execute postapproval commitments (e.g., epidemiologic studies, enhanced surveillance studies, other registries, REMS 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 in 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. However, there are several important limitations in interpreting spontaneous report data (also see Chapter 10). Because of the lack of complete numerator (number of cases) and the need to estimate the denominator (total number of patients actually exposed to the drug) data, 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 is 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.[31] 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 unbeknownst 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.[32] An example is the drug mibefradil, which was withdrawn from the market after less than a year by the manufacturer as a result of new information about multiple potentially serious drug interactions.[33] Adherence to medications also often differs between closely monitored trials and general postapproval use, as is the case with antihypertensives[34] (see also Chapter 42).

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.

The postapproval epidemiologic work conducted in support of sildenafil provides an example of how observational methods can be successfully used to provide additional scientific evidence regarding the safety of a newly approved medication. Sildenafil was approved in the US for the treatment of erectile dysfunction (ED) in March 1998, followed by an approval in the EU in May 1998. Immediately following the launch of sildenafil in the US, spontaneous reports of death and myocardial infarction among users of sildenafil were

received by the manufacturer and regulatory authorities. The volume of sildenafil spontaneous reports, in particular those from consumers, was unlike patterns seen for other new drugs at the time, and was unusual enough to raise regulatory concerns about its safety. Studies conducted prior to sildenafil's approval highlighted the prevalence of cardiovascular risk factors in patients with ED and evidence that ED can be an early warning sign of cardiovascular disease,[35] but the exact risk and predictors of acute cardiovascular events that occur among men with ED who seek and receive treatment were unknown at the time. Thus, in response to concerns raised by European regulators, two postapproval safety studies were initiated to investigate the postmarketing safety of sildenafil.

To obtain data on sildenafil's postmarketing safety in a timely manner, the first study undertaken was a two-stage UK Prescription Event Monitoring (PEM) study,[36] conducted by an independent academic center, the Drug Safety Research Unit (DSRU) at Southampton University in the UK (see Chapter 20). In this case, a PEM study was the only feasible data source by which results could be obtained rapidly, as it was not possible to use automated administrative or medical records databases since sildenafil was not reimbursed by these health systems. In all, more than 28 000 men were followed for a mean of 17.5 months.[37,38] The age-standardized mortality ratio in men using sildenafil, compared to the general English male population, indicated that mortality among sildenafil users was not elevated when compared to the 1998 rates in English men. These studies supported clinical trial evidence that the incidence of death due to cardiovascular disease among men receiving a prescription for sildenafil in a clinical practice setting is similar to the rate observed in men not using sildenafil.[38] Further, and most importantly, no cardiovascular or cerebrovascular safety signals were identified from the PEM study.

In addition to the UK PEM study, a prospective observational study, the International Men's Health Study (IMHS), was initiated to assess the occurrence of cardiovascular events in men receiving sildenafil for the treatment for ED. This cohort of 3813 men receiving prescriptions for sildenafil in Germany, France, Spain, and Sweden was followed for approximately 18 months on average to assess cardiovascular risk factors, cardiovascular events, and use of ED treatments.[39] The rates of cardiovascular disease events were found to be comparable to previously published figures from clinical trial and population-based epidemiologic data, providing further evidence supporting the cardiovascular safety of sildenafil.

These postapproval studies did not examine comparative safety, since sildenafil was the first in its class of drugs.[40] However, epidemiologic studies are often used to examine the comparative risks associated with particular drugs within a therapeutic class, as they are actually used in clinical practice. For example, one large study determined that among antiulcer drugs, cimetidine was associated with the highest risk of developing symptomatic acute liver disease.[41] Other studies, examining the risk of hip fractures in users of benzodiazepines, found that users of long-acting agents were at greater risk than those using short-acting agents.[42,43]

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 those actually exposed are at higher (or lower) risk of the adverse event than those unexposed, even in the absence of the medication. 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 37). 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.

As discussed above, the LST is a design used by epidemiologists when confounding is a large concern but real-world follow-up is critical (see Chapter 36). This was the approach adopted

for ziprasidone, an atypical antipsychotic for the treatment of schizophrenia launched in the US in 2001.[12] 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 comorbidities and risk factors. This possibility existed because ziprasidone would be a new atypical agent, and therefore most likely to be used initially among patients who had failed prior therapies. In addition, there were concerns 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.[44] Given these likely selection phenomena, random allocation of patients was the only approach providing the certainty of an unbiased comparison between groups. Thus, the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC) Large Simple Trial was designed to compare the safety of ziprasidone and olanzapine under real-world conditions (see Case study 7.1). ZODIAC demonstrated that there was no difference between the ziprasidone and olanzapine treatment arms with respect to non-suicide mortality (RR = 1.02, 95% CI: 0.79–1.39), and that 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).[13]

Drug safety evaluation in special populations

Pregnancy and birth outcomes

Unless a medication is being developed specifically to treat a pregnancy-related condition, pregnant women are generally excluded from clinical trials for ethical reasons, because of potential risks to the developing fetus and newborn.[45] 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. 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 (see Chapter 10). 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.[46] Observational follow-up methods are an effective way to monitor and assess the effects of medications on pregnant women and the fetus, in addition to birth outcomes and infant and child development.

Epidemiologic methods have been used to study cancers in individuals exposed to drugs *in utero*, periconceptually, or immediately after birth, and to examine possible teratologic effects of various agents (see Chapter 28). A classic example of exposure *in utero* is the association between maternal use of diethylstilbestrol and clear-cell adenocarcinoma of the vagina.[47,48] Other examples include the possible association between prenatal exposure to metronidazole and childhood cancer,[49] and childhood cancers and the use of sedatives during pregnancy.[50] A number of studies have examined the potential association between childhood cancer and exposure to vitamin K in the neonatal period.[51–54] Finally, although animal teratology testing is part of the preapproval process of all drugs, questions about a possible relationship between a specific drug and birth defects may arise in the postmarketing period. In these cases, epidemiologic methods are necessary to gather and evaluate the information in the population actually using the drug to examine possible teratogenicity. Such studies include those examining diazepam use and oral clefts;[55] spermicide use and Down's syndrome, hypospadias, and limb reduction deformities;[56] pyridoxine with doxylamine use and oral clefts, cardiac defects, and pyloric stenosis;[57,58] and antiviral drugs and birth defects.[59]

In certain circumstances, registries are used to obtain information about the safety of new medica-

tions during pregnancy (also see Chapter 21). The information provided by such registries allows health-care professionals and patients to make more informed choices on whether to continue or initiate drug use during pregnancy, or provides reassurance after a pregnancy has occurred on therapy, based on a benefit–risk analysis that can be conducted for each individual. The FDA and the European Medicines Agency (EMA) have issued guidelines on when it is appropriate to establish a pregnancy registry.[60,61]

A pregnancy exposure registry is typically prospective and observational, conducted to actively collect information about medication exposure during pregnancy and subsequent pregnancy outcome. Such registries differ from passive postmarketing surveillance systems in that they collect data from women prior to knowledge of the pregnancy outcome, proceeding forward in time from drug exposure to pregnancy outcome rather than backward in time from pregnancy outcome to drug exposure; this 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 will allow for a less biased measure of drug exposure 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. Finally, a pregnancy registry should also 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 criterion is ideally met 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 different treatment. In practice, however, it is usually not feasible to meet these criteria 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 followed.

In general, when analyzing data from pregnancy registries, those cases identified prospectively, (prior to knowledge of pregnancy outcomes) should be separated from those cases identified retrospectively, (after pregnancy outcome has been determined by prenatal diagnosis, abortion, or birth) as the latter will be biased towards reporting of abnormalities. To minimize ascertainment bias, incidence rates will ideally be calculated only from those cases identified prospectively. Also, since losses to follow-up may represent a higher proportion of normal pregnancy outcomes than abnormal pregnancy outcomes, participants in pregnancy registries should be intensively followed to obtain complete pregnancy outcome reports.

The FDA's Office of Women's Health is now maintaining a list of pregnancy exposure registries online,[62] including some for HIV/AIDS medications, the human papillomavirus and hepatitis B vaccines, and drugs for depression, migraines, diabetes, and other conditions.

Pregnancy registries may be sponsored by university-based research groups, by government agencies, by biopharmaceutical companies, or by collaborative efforts on the part of all three entities. While standard epidemiologic methods for estimating risks for pregnancy outcomes associated with drug exposures using data from pregnancy exposure registries have not yet been agreed upon, potential methods have been proposed and await further validation.[63]

Case–control studies are often considered more efficient for the study of rare outcomes, presenting the opportunity to test hypotheses generated by pregnancy registries. However, these studies tend to rely on maternal interviews after birth to establish data on drug exposure during pregnancy, which can introduce recall bias.

Large computerized health-care databases are increasingly being used to monitor exposure during

pregnancy with the development of maternal–infant record linkages. These databases, based on primary care data in Europe and health insurance claims in the US, represent large populations, giving studies substantial power. Internal control groups can be identified representing the general population and women with the same underlying condition exposed to no drug or other drugs within the class of interest. Although the bias associated with self-reported exposure data is avoided in these data sets, exposure data are based on prescription records which assume that the women took the medication close to the time the prescription was written or filled as instructed. Data on confounding variables are usually limited in these data sources.

Pediatrics

In the context of drug development, children are considered a special population. This categorization is due to the unique physiologic characteristics of children; developing organ systems often result in different and unpredictable pharmacokinetic and pharmacodynamic profiles from adults, beyond standard adjustments for smaller body size and weight (also see Chapter 3). 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 (also see Chapter 35).

There is drug development legislation specifically for pediatrics in both the EU and US. In the US in 2007 under FDAAA, two separate laws were renewed: the Best Pharmaceuticals for Children Act (BCPA) and the Pediatric Research Equity Act (PREA). The BCPA provides a voluntary incentive to study a drug in children by granting an additional 6 months of patent exclusivity on the moiety if adequate pharmacokinetic and efficacy–safety studies are conducted; this patent extension may be granted even if the drug was shown not to be effective or safe in children. Furthermore the company ultimately may choose not to market the drug even if it was approved for pediatric usage by the FDA. PREA, however, is mandatory and is limited to the indication under development. In addition, few exceptions are permitted except in cases where the disease does not exist in childhood (e.g., Alzheimer's Disease or postmenopausal osteoporosis), or when it is critical to first demonstrate adult safety and efficacy. The FDA may issue a written request for specific pediatric non-clinical and clinical studies, which can often be used to simultaneously satisfy the BCPA and PREA requirements.

The legislative revisions instituted in 2007 also created an internal Pediatric Review Committee (PeRC) within the FDA to better coordinate pediatric matters at the Agency, such as within the review division, but also to enhance transparency and communication of pediatric reviews and label changes with the public. The revisions include enhanced adverse event surveillance by requiring all spontaneous reports to be submitted with BCPA and PREA applications, as well as mandating public advisory committee review of pediatric adverse events. As BCPA and PREA were only renewed for 5 years, this legislation will expire in 2012; its renewal status is unknown as of publication.

In the EU, a substantial change to drug development for all biopharmaceutical companies resulted from the 2007 legislation that established the Paediatric Investigation Plan (PIP) process. The EU mandates that all new drugs, and any supplemental applications for new indications, even if not intended for pediatric use, submit a PIP no later than completion of adult pharmacokinetic studies (i.e., Phase I), or in advance of the supplemental application filing in the latter case.

Prior to the law, companies often did not begin pediatric study planning until the safety and efficacy of the drug had been established in adults and after at least one adult marketing authorization/approval was obtained. Like PREA, the new legislation allows very few exceptions, such as when the disease occurs in adults only. Similar to the EU-RMP, the PIP filing requires extensive information on the background epidemiology of the disease in children by age and other demographic and geographic subgroups, for which epidemiologists may play an important role in providing this information. The PIP process also requires that the adult RMP be expanded to include pediatric indications and to expand or conduct relevant safety studies, if applicable. This often results in separate pediatric observational safety studies as conditions of

approval, given the lack of long-term safety data that is typically available in this population.

Similar to the US, the EU legislation of 2007 also created a 6-month exclusivity incentive if the company satisfactorily fulfills its obligations outlined in the PIP, even if the drug was not shown to be safe or effective in children. It also created a pediatric scientific advice group, called the Pediatric Committee (PDCO), within the EMA to review the PIPs to assess the waiver and deferral requests, to assess whether a company has complied with elements agreed to in the PIP, and to review the data contained in the PIP for demonstration of adequate quality, efficacy, and safety of the product.

An important consequence of the PIP 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 lifecycle, must submit a PIP. The regulatory changes in the US and general attention towards enhanced 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 randomized clinical trial cannot be feasibly conducted, but also may require creative approaches to design and data source selection that would not be necessary in adult populations.

One example is in juvenile idiopathic (rheumatoid) arthritis (JIA). The incidence of JIA ranges from approximately 8 to 226 per 100 000 children,[64] whereas in US adults, rheumatoid arthritis (RA) is estimated to occur in approximately 1%,[65] or approximately 1000 per 100 000 population. Although there are explanations for some of these differences, a substantial disparity in incidence rates between adult and children remains.

The rarity of JIA versus RA highlights the need for creative long-term safety monitoring approaches. In the last decade, new treatments for both RA and JIA have emerged, such as the biologic tumor necrosis factor alpha (TNFα) inhibitors and T-cell costimulatory modulators, or traditional "small molecule" drugs such as the selective cyclooxygenase-2 (COX-2) inhibitors. These drugs were approved for adult RA based on testing programs that typically ranged in size from 2000 to 5000 patients;[66–68] for JIA, the largest program was less than 250 patients and were of substantially shorter duration.

The RA postmarketing safety programs in the US often utilize existing large insurance or other automated health-system databases to conduct efficient studies.[69] However for JIA, most biopharmaceutical companies have had to create independent multicenter prospective cohort studies or single-drug registries because the number of JIA patients exposed to the product of interest is very small and/or it would take many years to collect enough exposure in JIA patients to adequately assess rare outcomes in an automated database; further, the validity of many of these JIA diagnoses in automated databases would be suspect. By utilizing existing pediatric rheumatology networks and concentrating studies at centers that specialize in treating JIA patients, adequate sample sizes can be assembled more quickly and longer-term follow up is maximized because continuity of care is generally maintained even when insurance plans change (a limitation to which large insurance claims database studies are often prone). Maintaining long-term follow-up is critical in a growing population.

There are also novel enhanced surveillance and pharmacovigilance techniques that can be utilized to complement the information obtained from traditional observational studies. For example, a comprehensive risk management program for celecoxib use in JIA is comprised of a long-term observational registry, an enhanced surveillance program, a short-term randomized clinical trial, and formation of an external expert panel to review safety data from spontaneous reports and the three postapproval studies.[70–72] One of the unique aspects of this multifaceted program is the enhanced surveillance approach to identify all serious adverse events (SAEs) that occur in JIA patients, regardless of therapy, via a national network of pediatric rheumatologists. A very brief email survey is sent monthly to solicit possible cases and a simple form is used to capture critical information on any SAEs that have occurred. Sites also provide the total number of unique JIA patients seen to allow for calculation of more accurate reporting (incidence) rates than would normally be available using traditional spontaneous reporting rate estimates. Preliminary data suggest that the SAE rates identified in the simple JIA active surveillance program are comparable to those seen in other long-term observational studies of JIA patients.[73]

In Europe, the general lack of publicly available, large, automated data sources that contain specialist care information has stimulated a number of large, well-designed disease-based or drug-based registries to monitor the safety of biologics used in RA. These registries are unique because they are not a single-drug study supported by its manufacturer, but broad collaborations between government, professional rheumatologic societies, and multiple biopharmaceutical companies that have postmarketing commitments to monitor the long-term safety of their drugs. Similar JIA registries have emerged (e.g., the British Society of Pediatric and Adolescent Rheumatology in the UK or the JIA Register in Sweden) and some are beginning to link to the adult RA registries, allowing for follow-up well into adulthood. Of course, such registries raise issues of selection bias when they are not population based, especially if they do not have complete ascertainment of all patients in participating practices (see Chapter 21).

Rheumatology is not the only area where these methodologic, operational, and ethical issues arise; the study of pediatric oncology treatments faces many of the same concerns. In other instances, children are the primary target or vector for disease, as in many common infectious diseases. Adequate sample size may no longer be the primary methodologic limitation facing the epidemiologist, 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 26. Both situations highlight the need for novel approaches and methods 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 also makes significant contributions to the evaluation of the effectiveness of risk mitigation intervention measures (see also Chapter 29). In the last 5 years, this component of biopharmaceutical risk management has grown considerably. Guidances, such as the US FDA's Risk Minimization Action Plan (RiskMAP) issued in 2005 and associated with PDUFA III, outlined the tools industry could use to help reduce known or hypothetical risks when traditional minimization approaches (i.e., the product label) were insufficient.[74] These tools generally fall into three categories: enhanced education, for example patient labeling or prescriber training programs; reminder systems, for example patient consent forms or specialized packaging; or performance-linked access systems, for example requiring documentation of laboratory tests before each prescription or restricting distribution only to those who are certified prescribers. A critical addition to this guidance that was particularly relevant to epidemiologists within industry was the suggestion to perform assessments

of the effectiveness of these risk minimization tools and to submit these to the Agency for review.

The same year in Europe, Volume 9A of *The Rules Governing Medicinal Products in the European Union—Guidelines on Pharmacovigilance for Medicinal Products for Human Use* outlined the requirement to submit risk management plans for all new medicinal products, or new indications/dosage forms of existing products, or when requested by a competent authority. These risk management plans have an explicit section dedicated to "Ensuring the effectiveness of risk minimization activities/Assessment of risk minimization". Thus, in addition to requiring the filing of risk management plans, this legislation gives the EMA or any competent authority the right to require risk minimization activities such as restricted distribution or other evaluation of risk mitigation effectiveness.[75]

Risk evaluation and mitigation strategies (REMS)

Under FDAAA, the FDA can require a sponsor to submit a proposed risk evaluation and mitigation strategies (REMS) as part of its initial application if the FDA finds that a REMS is necessary to ensure the benefits of the drug or biological product outweigh the risks. The FDA may also require a REMS postapproval based upon "new safety information." FDAAA has defined this as any information obtained during the initial review process, as a result of postapproval studies, or spontaneous reports.[76] The REMS requirement is an expansion of RiskMAPs described in the *Risk Management Guidances* issued by the FDA in 2005. Conceptually the tools have remained similar, but the emphasis has shifted to medication guides, the mandatory patient information that is required to be distributed with each prescription for drugs under a REMS. Medication guides are intended to directly inform patients about the most important risks in lay language, in contrast to the lengthy and comprehensive information contained in product labels intended for prescribers. In addition, REMS can include communication plans or elements to assure safe use (ETASU), which correspond to the activities utilized in the "performance-linked access systems."

When determining whether to require a REMS, the REMS statute states that the FDA must consider the estimated size of the target use population, the seriousness of the treated condition, the expected treatment benefit, the treatment duration, the severity of known or potential adverse events, and whether or not the drug is a new molecular entity. All drugs and biologics may be eligible for REMS designations. Sponsor–Agency agreement on the necessity and scope of potential REMS is an element to be resolved during the drug or REMS approval process. Moreover, the possibility of a REMS remains throughout a product's lifecycle, and the FDA may impose civil monetary penalties for violations or non-compliance.

In September 2009, the FDA published a draft guidance for REMS, providing a framework for REMS submissions.[77] All REMS for patent-protected products must include a timetable for submission of the REMS assessments, typically at 18 months, 3 years, and 7 years following REMS approval. FDA may waive the 7-year requirement. Currently, the vast majority (~75%) of REMS only require a medication guide and assessment.[78] Medication guides are thus viewed as the primary tool in enhanced risk mitigation efforts, despite several studies suggesting dubious effectiveness.[79,80] The results of REMS assessments should provide further important insight to this much-needed area of research.

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 health-care or claims database, assessments may measure compliance with monitoring guidelines or measure 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, and require the epidemiologist to design cross-sectional surveys specific for each recipient, drug, and associated unique risk profile, given that standardized or

validated questionnaires that measure drug-specific medication guide comprehension do not exist. An example of a comprehensive natalizumab REMS program is shown in Case study 7.3.

The implementation of the REMS and EU-RMP legislation has highlighted a number of difficulties. The mandated assessment timelines associated with REMS may be difficult to achieve for many reasons: the need to develop and pilot knowledge/comprehension surveys unique to each drug covered under a REMS; to design, implement, and assess complex safe use programs; the scarcity of patients treated with the drug of interest; or difficulties in identifying them through automated channels. The fractured health-care and prescription delivery system in the United States presents a barrier to efficient distribution of medication guides and educational materials, and certainly to the implementation of many safe use elements.

In addition, as of publication, there is no FDA or EMA guidance that provides detailed information on the preferred methods to assess the effectiveness of risk mitigation activities. There is little information in the peer-reviewed literature to provide a scientific basis for the utility of medication guides, or to provide information on what constitutes an "effective" risk mitigation program, on what constitutes an "important" or "meaningful" change in knowledge/comprehension, or what is the minimally acceptable level of comprehension.[85–87] The lack of regulatory or scientific guidance currently leaves the industry and regulators in a position of exploration and building iterative knowledge about the preferred methods for behavioral risk intervention, including how they vary across and within patient populations. Better information in this area is especially critical when the ideal measurement of the true effectiveness of a risk mitigation program—a decrease in or elimination of the adverse event of interest associated with a drug—may be difficult to measure, or infer that it is a result of the intervention. Knowledge in these areas is expected to mature as more companies and the regulatory agencies garner additional experience.

Risk mitigation evaluation is thus an emerging area for epidemiologists in industry. Specialized expertise in survey design and implementation, observational study experience using both primary and secondary data collection methods, program and behavioral risk intervention evaluation, and data analysis are clearly critical to successful evaluation of risk mitigation activities now required by legislation in the US and EU.

Collaborations in research efforts and drug safety initiatives

Pharmacoepidemiology is a constantly evolving field, with changes in areas such as pharmacogenomics (Chapter 34) and comparative effectiveness research (Chapter 32) occurring rapidly; future directions in the field are also covered in the final chapter of this book (Chapter 48). Nonetheless, there are some emerging topics that are important specifically for those involved in or with industry. The most important is the increased collaboration across all disciplines, including collaboration between biopharmaceutical companies to further pharmacoepidemiologic research approaches and sources, and to enhance the study of drug safety in general. The goal of these cross-sector collaborations is to combine data to increase scientific/logistical efficiency and sample size, and to pool scarce resources. These collaborations tend to be either (i) disease area- or subpopulation-specific, or (ii) broad drug safety initiatives.

In the sphere of disease area- or subpopulation-specific collaborations, there are several ongoing and successful examples. One is the Highly Active Antiretroviral Therapy Oversight Committee (HAART OC), comprised of manufacturers of antiretroviral medications for the treatment of HIV/AIDS, US and European regulatory agencies, academics, and patient advocates. HAART OC has collectively sponsored observational studies to fulfill multicompany postapproval commitments to assess the risk of cardiovascular morbidity and mortality of these drugs, as well as to construct validated case definitions of clinical phenomena such as lipodystrophy.[88–90] Similar consortia exist to study the risk of birth defects potentially associated with antiepileptic drug (AED) use during pregnancy in North America (the North American AED Pregnancy Registry) and Internationally (EURAP, the European

Case study 7.3 Risk management: natalizumab REMS

The FDA approved natalizumab, the first humanized monoclonal antibody for the treatment of relapsing multiple sclerosis (MS) via accelerated approval in November 2004. Approximately 3 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.[81] A third fatal case was identified in a Crohn's disease patient shortly thereafter.[82]

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 immunocompetant 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.[81]

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 (Figure 7.1).[83,84] The goals of TOUCH™ are as follows.

Risk assessment goals:

- determine the incidence and risk factors for PML
- assess long-term safety in clinical practice.

Risk minimization goals:

- promote informed benefit–risk decisions
- minimize the risk of PML
- potentially minimize death and disability.

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/IDEC 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.[82,84]

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.

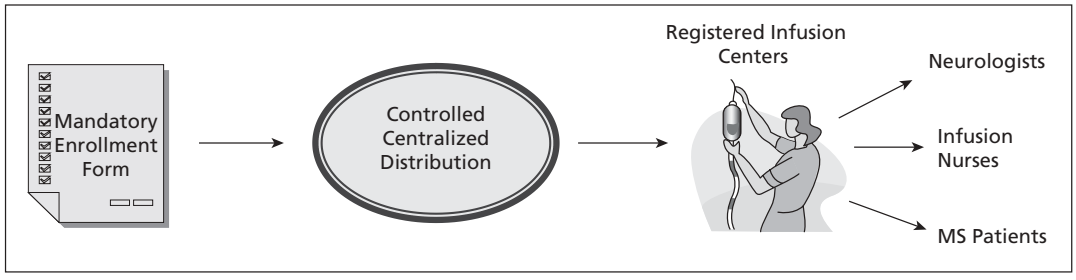


Figure 7.1 TOUCH risk minimization system.

Registry of AEDs in Pregnancy, which now includes countries beyond Europe).[91,92]

Alternatively, some cross-sector partnerships are disease-based rather than drug-based, such as the new JIA CORE (US) or PharmaChild (EU) initiatives to better characterize the safety profile of new and existing immunomodulatory drugs used in JIA;[93,94] or the well-established Cystic Fibrosis Foundation Patient Registry in the US which has been continuously collecting data for over 40 years.[95]

While great efficiency is gained by combining efforts, all approaches present unique logistical, legal, ethical, and regulatory challenges. The structure and governance of these consortia demand flexibility to meet various scientific and regulatory needs to fulfill an individual company's postapproval study requirements. Furthermore, as these are long-term endeavors, they require adaptability and scalability since safety questions and treatment paradigms often shift considerably over time.

In addition to disease-area-specific collaborations, epidemiologists in the biopharmaceutical industry are active contributors to drug safety initiatives in the US and EU designed to advance the field of pharmacoepidemiology, including the Observational Medical Outcomes Partnership (OMOP)[96] and the FDA Sentinel Initiative (see Chapter 30) in the US; the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP),[97] and the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT), one of the projects of Innovative Medicine Initiative (IMI) in Europe.[98]

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 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, the need for scientists with training and research experience in pharmacoepidemiology has never been greater. To best support drug safety evaluation epidemiologic strategies must: (i) begin early in development, (ii) continue throughout the lifecycle of the drug, (iii) evolve as new safety information becomes available, and (iv) be innovative, requiring epidemiologists to be aware of new methods and methods specific to the disease area. Epidemiologists within industry have an opportunity to build on the successes of the last 30 years by collaborating with academics, non-profit organizations, and regulators to advance the methods of drug safety evaluation and risk management.

References

1. Berger MS, Berger BA. *FDA Overview*. http://www.emedicinehealth.com/fda_overview/article_em.htm (accessed June 2011).
2. Harris G. *What's Behind an F.D.A. Stamp?* 2008. http://www.nytimes.com/2008/09/30/health/policy/30fda.html?_r=2&adxnnl=1&adxnnlx=1282766444-+QZN1bSKTwSlQ54i7CIceQ (accessed June 2011).
3. Institute of Medicine. *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 2006. http://www.iom.edu/Reports/2006/The-Future-of-Drug-Safety-Promoting-and-Protecting-the-Health-of-the-Public.aspx (accessed June 2011).
4. Food and Drug Administration. *Guidance for Industry Postmarketing Studies and Clinical Trials*, 2011. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM172001.pdf (accessed August 2011).
5. International Society for Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**: 200–8.
6. FDA. Center for Drug Evaluation and Research. Psychopharmacologic Drugs Advisory Committee, 2000. http://www.fda.gov/ohrms/dockets/ac/00/transcripts/3619t1.rtf (accessed June 2011).

7. Harrigan EP, Miceli JJ, Anziano R, Watsky E, Reeves KR, Cutler NR, *et al*. A randomized evaluation of the effects of six antipsychotic agents on QTc, in the absence and presence of metabolic inhibition. *J Clin Psychopharmacol* 2004; **24**: 62–9.
8. Enger C, Weatherby L, Reynolds RF, Glasser DB, Walker AM. Serious cardiovascular events and mortality among patients with schizophrenia. *J Nerv Ment Dis* 2004; **192**: 19–27.
9. Curkendall SM, Mo J, Glasser DB, Rose Stang M, Jones JK. Cardiovascular disease in patients with schizophrenia in Saskatchewan, Canada. *J Clin Psychiatry* 2004; **65**: 715–20.
10. Hennessy S, Bilker WB, Knauss JS, Margolis DJ, Kimmel SE, Reynolds RF, *et al*. Cardiac arrest and ventricular arrhythmia in patients taking antipsychotic drugs: cohort study using administrative data. *BMJ* 2002; **325** (7372): 1070.
11. Hennessy S, Bilker WB, Knauss JS, Kimmel SE, Margolis DJ, Morrison MF, *et al*. Comparative cardiac safety of low-dose thioridazine and low-dose haloperidol. *Br J Clin Pharmacol* 2004; **58**: 81–7.
12. Strom BL, Faich GA, Reynolds RF, Eng SM, D'Agostino RB, Ruskin JN, *et al*. The Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC): design and baseline subject characteristics. *J Clin Psychiatry* 2008; **69**: 114–21.
13. Strom BL, Eng SM, Faich G, Reynolds RF, D'Agostino RB, Ruskin J, *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).appi. *Am J Psychiatry* 2011; **168**: 193–201.
14. Campbell UB, Petronis KR, Klee BJ, Kolitsopoulos F, Walker AM, Mo J. Case-crossover study of PDE5 inhibitor exposure as a potential "trigger factor" for acute NAION. *Pharmacoepidemiol Drug Saf* 2009; **18** (S1): S166–7.
15. 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**: 541–9.
16. Melfi CA, Croghan TW. Use of claims data for research on treatment and outcomes of depression care. *Med Care* 1999; **37**: AS77–80.
17. 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**: 211–8.
18. 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**: 9–15.
19. Hennessey S, Bilker WB, Weber A, Strom BL. Descriptive analyses of the integrity of a US Medicaid claims database. *Pharmacoepidemiol Drug Saf* 2003; **12**: 103–11.
20. Eng P, Seeger JD, Loughlin J, Oh K, Walker AM. 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**: 101–7.
21. 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.
22. Eye Diseases Prevalence Research Group. Causes and prevalence of visual impairment among adults in the United States. *Arch Ophthalmol* 2004; **122**: 477–85.
23. Gragoudas ES, Adamis AP, Cunningham ET, Jr., Feinsod M, Guyer DR. Pegaptanib for neovascular age-related macular degeneration. *N Engl J Med* 2004; **351**: 2805–16.
24. Mo J, Duan Y, Patel M, Klein R, Scott IU, Huang K, *et al*. Risk of intravitreal injection-related endophthalmitis in patients with age-related macular degeneration (AMD) in the US Medicare population. *Pharmacoepidemiol Drug Saf* 2006; **15** (S1): S215–6.
25. Liao D, Mo J, Duan Y, Huang K, Klein R, Scott IU, *et al*. Incidence of intravitreal injection (IVT) related endophthalmitis among AMD patients. *Annual Meeting of American Academy of Ophthalmology and European Society of Ophthalmology*. Atlanta, Georgia, 2008.
26. Aiello LP, Brucker AJ, Chang S, Cunningham ETJR, D'Amico DJ, Flynn HWJ, *et al*. Evolving guidelines for intravitreous injections. *Retina* 2004; **24** (Suppl): S3–S19.
27. 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.
28. 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**: 29–36.
29. Derby LE, Tennis P, Jick H. Sudden unexplained death among subjects with refractory epilepsy. *Epilepsia* 1996; **37**: 931–5.

30. Leestma JE, Annegers JF, Brodie MJ, Brown S, Schraeder P, Siscovick D, *et al*. Sudden unexplained death in epilepsy: observations from a large clinical development program. *Epilepsia* 1997; **38**: 47–55.
31. FDA. *From Test Tube to Patient: Improving Health Through Human Drugs*. Washington, DC: US Food and Drug Administration, 1999.
32. Rogers AS. Adverse drug events: identification and attribution. *Drug Intell Clin Pharm* 1987; **21**: 915–20.
33. FDA. *Roche Laboratories Announces Withdrawal of Posicor from the Market*. FDA Talk Paper, June 8 1998.
34. Andrade SE, Walker AM, Gottlieb LK, Hollenberg NK, Testa MA, Saperia GM, *et al*. Discontinuation of antihyperlipidemic drugs—do rates reported in clinical trials reflect rates in primary care settings? *N Engl J Med* 1995; **332**: 1125–31.
35. Jackson G. Erectile dysfunction and cardiovascular disease. *Int J Clin Prac* 1999; **53**: 363–8.
36. Mann RD. Prescription-event monitoring—recent progress and future horizons. *Br J Clin Pharmacol* 1998; **46**: 195–201.
37. Shakir SA, Wilton LV, Boshier A, Layton D, Heeley E. Cardiovascular events in users of sildenafil: results from first phase of prescription event monitoring in England. *BMJ* 2001; **322**: 651–2.
38. Boshier A, Wilton LV, Shakir SA. Evaluation of the safety of sildenafil for male erectile dysfunction: experience gained in general practice use in England in 1999. *Br J Urol* 2004; **93**: 796–801.
39. Mittleman MA, Maclure M, Lewis MA, Hall GC, Moore N, Giuliano F, *et al*. Cardiovascular outcomes among sildenafil users: results of the International Men's Health Study. *Int J Clin Prac* 2008; **62**: 367–73.
40. 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**: 1861–8.
41. Garcia Rodriguez LA, Wallander MA, Stricker BH. The risk of acute liver injury associated with cimetidine and other acid-suppressing anti-ulcer drugs. *Br J Clin Pharmacol* 1997; **43**: 183–8.
42. Ray WA, Griffin MR, Downey W. Benzodiazepines of long and short elimination half-life and the risk of hip fracture. *JAMA* 1989; **262**: 3303–7.
43. Cummings SR, Nevitt MC, Browner WS, Stone K, Fox KM, Ensrud KE, *et al*. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. *N Engl J Med* 1995; **332**: 767–73.
44. Komossa K, Rummel-Kluge C, Schmid F, Hunger H, Schwarz S, El-Sayeh HG, *et al*. Aripiprazole versus other atypical antipsychotics for schizophrenia. *Cochrane Database Syst Rev* 2009; **4**: CD006569.
45. 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.
46. Henshaw SK. Unintended pregnancy in the United States. *Fam Plann Perspect* 1998; **30**: 24–9.
47. 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.
48. Herbst AL, Kurman RJ, Scully RE, Poskanzer DC. Clear-cell adenocarcinoma of the genital tract in young females. Registry report. *N Engl J Med* 1972; **287**: 1259–64.
49. Thapa PB, Whitlock JA, Brockman Worrell KG, Gideon P, Mitchel EF, Jr., Roberson P, *et al*. Prenatal exposure to metronidazole and risk of childhood cancer: a retrospective cohort study of children younger than 5 years. *Cancer* 1998; **83**: 1461–8.
50. Kinnier-Wilson LM, Kneale GW, Stewart AM. Childhood cancer and pregnancy drugs. *Lancet* 1981; **2** (8241): 314–15.
51. Olsen JH, Hertz H, Blinkenberg K, Verder H. Vitamin K regimens and incidence of childhood cancer in Denmark. *BMJ* 1994; **308**: 895–6.
52. Parker L, Cole M, Craft AW, Hey EN. Neonatal vitamin K administration and childhood cancer in the north of England: retrospective case–control study. *BMJ* 1998; **316**: 189–93.
53. McKinney PA, Juszczak E, Findlay E, Smith K. Case–control study of childhood leukaemia and cancer in Scotland: findings for neonatal intramuscular vitamin K. *BMJ* 1998; **316**: 173–7.
54. von Kries R, Gobel U, Hachmeister A, Kaletsch U, Michaelis J. Vitamin K and childhood cancer: a population based case–control study in Lower Saxony, Germany. *BMJ* 1996; **313**: 199–203.
55. Rosenberg L, Mitchell AA, Parsells JL, Pashayan H, Louik C, Shapiro S. Lack of relation of oral clefts to diazepam use during pregnancy. *N Engl J Med* 1983; **309**: 1282–5.
56. Louik C, Mitchell AA, Werler MM, Hanson JW, Shapiro S. Maternal exposure to spermicides in relation to certain birth defects. *N Engl J Med* 1987; **317**: 474–8.
57. 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**: 2311–14.
58. Mitchell AA, Schwingl PJ, Rosenberg L, Louik C, Shapiro S. Birth defects in relation to Bendectin use

in pregnancy. II. Pyloric stenosis. *Am J Obstet Gynecol* 1983; **147**: 737–42.

59. Pasternak B, Hviid A. Use of acyclovir, valacyclovir, and famciclovir in the first trimester of pregnancy and the risk of birth defects. *JAMA* 2010; **304**: 859–66.
60. Food and Drug Administration. *Guidance for Industry: Establishing Pregnancy Exposure Registries*, 2002. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071639.pdf (accessed June 2011).
61. 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 June 2011).
62. FDA. *List of Pregnancy Exposure Registries*. http://www.fda.gov/ScienceResearch/SpecialTopics/WomensHealthResearch/ucm134848.htm (accessed June 2011).
63. Goldstein DJ, Sundell KL, DeBrota DJ, Offen WW. Determination of pregnancy outcome risk rates after exposure to an intervention. *Clin Pharmacol Ther* 2001; **69**: 7–13.
64. Manners PJ, Bower C. Worldwide prevalence of juvenile arthritis why does it vary so much? *J Rheumatol* 2002; **29**: 1520–30.
65. Shapira Y, Agmon-Levin N, Shoenfeld Y. Geoepidemiology of autoimmune rheumatic diseases. *Nat Rev Rheumatol* 2010; **6**: 468–76.
66. FDA. *Celecoxib*, page 10. http://www.accessdata.fda.gov/drugsatfda_docs/nda/98/20998AP_medr_P8.pdf (accessed July 2011).
67. FDA. *FDA Advisory Committee Briefing Document. NDA 21-042, s007VIOXX Gastrointestinal Safety*, 2001, page 5. http://www.fda.gov/ohrms/dockets/ac/01/briefing/3677b2_03_med.pdf (accessed July 2011).
68. FDA. *FDA Advisory Committee Meeting. Cardiovascular Safety of Valdecoxib and Parecoxib and Risk-Benefit Assessment*, 2005, slide #7. http://www.fda.gov/ohrms/dockets/ac/05/slides/2005-4090S1_10_Pfizer-Valdecoxib-Parecoxib_files/frame.htm (accessed July 2011).
69. FDA. *Postmarket Requirements and Commitments*. http://www.accessdata.fda.gov/scripts/cder/pmc/index.cfm (accessed June 2011).
70. Sobel RE, Lovell DJ, Giannini EH, Sheth N, Abramsky S, Cawkwell G. SINCERE™: A new multicenter registry to monitor long-term safety of non-steroidal anti-inflammatory drugs (NSAIDs) including celebrex in juvenile idiopathic arthritis (JIA). *Pharmacoepidemiol Drug Saf* 2008; **17** (Suppl 1): S265–6.
71. Sobel RE, Lovell DJ, Giannini EH, Sheth N, Abramsky S, Cawkwell G. A new multi-center registry to monitor long-term safety of non-steroidal anti-inflammatory drugs (NSAIDs) including celebrex in juvenile idiopathic arthritis (JIA). *Proceedings of the 14th Pediatric Rheumatology European Society Rheumatology Congress*, 2007, p. 193.
72. Sheth N, Sobel RE, Cawkwell G, West C. A novel and proactive approach to evaluate the safety of NSAIDs in juvenile idiopathic arthritis (JIA) patients. *Proceedings of the 14th Pediatric Rheumatology European Society Rheumatology Congress*, 2007, p. 184.
73. Wallace C, Hendrickson A, Sobel RE. Enhanced drug safety surveillance (EDSS) pilot project in JIA. *Arthritis Rheum* 2009; **60** (10 Suppl): S91.
74. FDA. *Guidance for Industry: Development and Use of Risk Minimization Action Plans*. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071616.pdf (accessed June 2011).
75. Committee for Medicinal Products for Human Use (CHMP). *Volume 9A of The Rules Governing Medicinal Products in the European Union Guideline on Pharmacovigilance for Medicinal Products for Human Use*, 2008. http://ec.europa.eu/health/files/eudralex/vol-9/pdf/vol9a_09-2008_en.pdf (accessed July 2011).
76. FDA. *Food and Drug Administration Amendments Act of 2007*. http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/SignificantAmendmentstotheFDCAct/FoodandDrugAdministrationAmendmentsActof2007/FullTextofFDAAALaw/default.htm (accessed June 2011).
77. FDA. *Draft Guidance for Industry: Format and Content of Proposed Risk Evaluation and Mitigation Strategies (REMS), REMS Assessments, and Proposed REMS Modifications*, 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM184128.pdf (accessed June 2011).
78. FDA. *Approved Risk Evaluation and Mitigation Strategies (REMS)*. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111350.htm (accessed June 2011).
79. Wolf MS, Davis TC, Shrank WH, Neuberger M, Parker RM. A critical review of FDA-approved Medication Guides. *Patient Educ Couns* 2006; **62**: 316–22.
80. Allen LaPointe N, Pappas P, Deverka P, Anstrom K. Patient receipt and understanding of written information provided with isotretinoin and estrogen prescriptions. *J Gen Intern Med* 2007; **22**: 98–101.

81. FDA. *Natalizumab (Marketed as Tysabri) Questions and Answers*, 2005. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm107202.htm (accessed June 2011).
82. FDA. *Information on Natalizumab (marketed as Tysabri)*. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm107198.htm (accessed June 2011).
83. *The TYSABRI TOUCH® Prescribing Program*. http://www.tysabri.com/tysbProject/tysb.portal/_baseurl/threeColLayout/SCSRepository/en_US/tysb/home/treatment-with-tysabri/touch-prescribing-program.xml (accessed June 2011).
84. Bozic C. Can a risk management program save a product from withdrawal? *46th Annual Drug Information Association Meeting*, 2010.
85. Adams EH. *REMS: A Year Later, Real World Experience*, 2009. http://www.thci.org/Opioid/jul09Docs/adams.pdf (accessed June 2011).
86. *Patient Safety Implications on Implementation of the Current FDA-Mandated Medication Guide Program*, 2006. http://www.healthcaredistribution.org/gov_affairs/pdf_addgovagency/NACDSMedGuide_positionpaper.pdf (accessed June 2011).
87. McEvoy GK. *Use of Medication Guides to Distribute Drug Risk Information to Patients*, 2007. http://www.fda.gov/downloads/Drugs/DrugSafety/UCM173475.pdf (accessed June 2011).
88. Friis-Moller N, Sabin CA, Weber R, d'Arminio Monforte A, El-Sadr WM, Reiss P, *et al*. Combination antiretroviral therapy and the risk of myocardial infarction. *N Engl J Med* 2003; **349**: 1993–2003. [Erratum appears in *N Engl J Med* 2004; **350**: 955.]
89. Worm SW, Sabin C, Weber R, Reiss P, El-Sadr W, Dabis F, *et al.* Risk of myocardial infarction in patients with HIV infection exposed to specific individual antiretroviral drugs from the 3 major drug classes: the data collection on adverse events of anti-HIV drugs (D: A: D) study. *J Infect Dis* 2010; **201**: 318–30.
90. Monforte A, Abrams D, Pradier C, Weber R, Reiss P, Bonnet F, *et al*. HIV-induced immunodeficiency and mortality from AIDS-defining and non-AIDS-defining malignancies. *AIDS* 2008; **22**: 2143–53.
91. Holmes LB, Wyszynski DF, Lieberman E. The AED (antiepileptic drug) pregnancy registry: a 6-year experience. *Arch Neurol* 2004; **61**: 673–8.
92. Tomson T, Battino D, Craig J, Hernandez-Diaz S, Holmes LB, Lindhout D, *et al*. Pregnancy registries: differences, similarities, and possible harmonization. *Epilepsia* 2010; **51**: 909–15.
93. 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.
94. *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).
95. Cystic Fibrosis Foundation. *Patient Registry Report*. http://www.cff.org/research/ClinicalResearch/PatientRegistryReport/ (accessed June 2011).
96. Observational Medical Outcomes Partnership. http://omop.fnih.org/ (accessed June 2011).
97. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). http://www.encepp.eu/ (accessed June 2011).
98. *Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium*. http://www.imi-protect.eu/ (accessed June 2011).

CHAPTER 8
The Role of Pharmacoepidemiology in Regulatory Agencies

Gerald J. Dal Pan[1] *and Peter Arlett*[2]
[1] Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA
[2] Pharmacovigilance and Risk Management, European Medicines Agency, London, UK

Introduction

The role of regulatory agencies in the regulation of medicines is broad, spanning the life cycle of a medicine, from the first human studies through the entire marketing period. While the range of individual activities is wide, the fundamental purposes of drug regulation include overseeing clinical research and the protection of human subjects in the prelicensing phase, granting access to medicines via licensing, monitoring the safety of medicines in the postlicensing phase, monitoring pharmaceutical advertising and promotion in the postlicensing phase, and insuring the quality of medicines. The regulator's approach is one with a public-health focus, based in science, and executed in the context of applicable laws and regulations. Within that framework, pharmacoepidemiology is playing a growing role.

At its core, pharmacoepidemiology seeks to describe the use and effects of medicines in a population. For a drug regulator, this is clearly a relevant science, one which contributes robust data upon which to make sound decisions regarding licensing, postlicensing safety monitoring, and, increasingly, in managing the known risks of marketed medicines.

For the regulators, three aspects of pharmacoepidemiology are particularly important at this time. First, as the scope of pharmacoepidemiology broadens, it is used increasingly throughout the lifecycle of a medicine. The roles of pharmacoepidemiology vary across the lifecycle, but at all times seek to understand the impact of medicines and their use in the population.

Second, the synthesis of data from multiple sources, many of which may rely heavily on pharmacoepidemiology, is critical for sound regulatory decisions. Synthesizing the data is challenging, but it is critical for the regulator to weigh all sources of evidence and arrive at a regulatory action that is based on a clear and transparent integration of all available data.

Third, building capacity and collaboration in pharmacoepidemiology is essential, both within and outside of regulatory agencies, in order for regulators, industry, and academia to meet the demand for high-quality pharmacoepidemiologic investigations.

The scope of pharmacoepidemiology throughout the product lifecycle

In the past two decades, the role of pharmacoepidemiology, from the regulator's perspective, has

The views expressed herein are those of the authors, and not necessarily of the US Food and Drug Administration or the European Medicines Agency.

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

grown beyond postapproval risk assessment to encompass assessing the need for medicines, planning certain aspects of drug development programs, evaluating preapproval clinical safety data, planning postapproval safety studies, monitoring postapproval safety, assessing actual use patterns of a medicine, and measuring the impact of regulatory actions. Each of these aspects is discussed in more detail below.

Assessing the need for medicines

Pharmacoepidemiology, and clinical epidemiology more broadly, can be used in drug development long before a medicine is licensed or even tested in humans. Population-based databases can be used to characterize the frequency and distribution of specific diseases, so that relevant populations can be included in the developmental clinical trials. Health-care 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 potential adverse events that arise during development in context. 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, is a well known methodologic problem in non-randomized pharmacoepidemiologic studies (see also Chapters 3 and 37), 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.

Characterizing the frequency of specific diseases can also be important in the development of medicines for rare diseases. For example, orphan drug programs are designed to provide incentives to pharmaceutical manufacturers who develop medicines for rare conditions, known as "orphan drugs." In the United States, an orphan drug designation is given to a drug or biologic that has shown promise as a therapy intended to treat disease affecting fewer than 200 000 persons in the United States.[1] In the European Union (EU), a prevalence of five per 10 000 persons in the EU is used.[2] 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.[3] (A less widely used provision of the orphan drug regulations in the US designates orphan drug status to a drug intended for diseases or conditions affecting 200 000 or more people if there is no reasonable expectation that costs of research and development of the drug for the indication can be recovered by sales of the drug in the United States. A similar, "insufficient return on investment" clause exists in the EU Orphan Regulation and is also little used.)

Pharmacoepidemiology is central to the designation of a product as an orphan drug product, as determination of prevalence is the basis for such designation. Methods for determining prevalence can include administrative health-care databases, electronic medical record systems, registries, and surveys. Most of these methods would 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. For estimates of population prevalence near the threshold, care must be taken to ensure that the most rigorous methods have been used. In this setting, regulators must ensure that the prevalence of the condition or disease does not exceed the threshold. The closer the estimated prevalence is to the threshold, the greater the precision needed to characterize the prevalence.

A review of 25 years' experience with the orphan drug program in the United States indicated that 1892 orphan designations had been granted. The median prevalence of the condition being treated was 39 000; the most common patient prevalence was 10 000 or fewer patients. Relatively few prevalence estimates were near the 200 000 threshold.[1]

Planning drug development programs

Regulators understand that there are not adequate therapies for certain serious or life-threatening diseases, and that development programs that require definitive evidence of an effect on irreversible morbidity or mortality may be very long and delay

access of effective therapies to patients. To allow patient access as rapidly as is feasible, and to assure that definitive evidence of effectiveness is obtained, the concept of "accelerated approval" has been developed. Under this framework, the FDA may grant marketing approval for a new drug product on the basis of adequate and well-controlled clinical trials establishing that the drug product has an effect on a surrogate endpoint that is reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit, or on the basis of an effect on a clinical endpoint other than survival or irreversible morbidity.[4] 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 1-year time-limited authorization where the authorizations renewal is linked to further data submission.[5] 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 as to the relation 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.

Understanding the relationship between a surrogate endpoint and a disease outcome is an opportunity for pharmacoepidemiologists to contribute to drug development. The hallmark of a surrogate marker is that it is reasonably likely to predict a clinical outcome of interest, even if it is not itself a direct measure of that clinical outcome. For example, in the field of oncology, improved overall survival is an outcome of clinical interest.[6] Although this outcome measure can be reliably measured in clinical trials, it can take a long time to generate the data needed to demonstrate an improvement in overall survival. To accelerate drug development, alternative outcome measures, which do not require such lengthy trials but which are believed to predict overall survival, can be used. Such surrogate measures could include disease-free survival, objective response rate, complete response rate, and progression free survival. However, these measures have not been validated as surrogates for overall survival in all settings. In addition, these outcomes are less precisely measured than overall survival. The pharmacoepidemiologist can play a role in establishing the relationship between this type of surrogate marker and a clinical outcome of interest.

Preapproval review of clinical safety data

The traditional role of pharmacoepidemiology, from a regulatory standpoint, has been the assessment of the safety of medicines in the postlicensing period. The limitations of prelicensing clinical trials in defining the full scope of adverse drug reactions are well known. Clinical trials are relatively small in size, compared to the size of the population of patients who 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 few data, or no data at all, in certain populations such as children, the elderly, and pregnant women. These groups, however, are often treated with the medicines in the course of clinical practice.

The analytic methods of clinical trials are best suited for data arising from randomized, 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. The same approach can be applied to protocols designed

to grant expanded access to a medicine before it is approved, such as may occur during a "treatment protocol" in the United States. Such protocols are typically single-arm, open-label studies.

Planning for postapproval safety studies

At the time a medicine is approved, it is well known that there are uncertainties and unknowns regarding the safety profile of the medicine. 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.

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.[7] The strategy starts prior to approval, when data from clinical trials are examined to determine the cardiovascular risk of the new medicine relative to that of comparative agents. A relative risk estimate is calculated. If the upper limit of the 95% confidence limits 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 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 postlicensing assessment of the safety of medicines involves both a proactive approach and, of necessity, a reactive approach. Proactive strategies involve carefully identifying important gaps in the safety profile of a medicine, and designing observational studies or clinical trials to address the unanswered questions. The approach to studying the cardiovascular risk of antidiabetic agents, noted above, is an example of a proactive step taken at the time of approval. However, the identification of knowledge gaps can come anywhere 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, working with industry, to develop a proactive approach to drug safety issues in the postapproval period. In the EU, the planning of postapproval studies is formalized as the Pharmacovigilance Plan, part of the EU Risk Management Plan where identification of risks and of knowledge gaps is based on the Safety Specification.[8]

Reactive approaches are also needed in regulatory pharmacoepidemiology because the adverse effects of medicines can become recognized at any time, including several years, after approval. To the extent that regulators can use proactive pharmacoepidemiologic approaches, reactive approaches can be minimized. But 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 a new safety issue is identified from spontaneously reported suspected adverse drug reactions (see Chapter 10), or when drug safety findings are published by independent groups, and neither the regulator nor the manufacturer is aware of them. 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.[9]

The specific scientific approach to an individual postapproval safety issue is beyond the scope of

this chapter. From the regulator's point of view, scientific undertakings to address new safety issues in the postapproval setting must be designed in ways that address the specific questions that regulators have, so that the results of these scientific undertakings can appropriately inform regulatory actions. As a regulator will use scientific information to reach sometimes urgent regulatory decisions that impact directly on the use of a medicine and therefore on product safety, the scientific investigations must be those that answer the question as accurately as possible in as little time as possible.

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 a medicine is actually used in clinical practice. Understanding the actual use of a medicine in practice allows regulators to assess the degree to which the medicine is used in ways that are consistent with the safe use of the medicine. To do so, regulators can use a variety of pharmacoepidemiologic resources, including administrative claims data, electronic medical records, or other public-health databases.

For example, Kaplan and colleagues examined the use patterns of metoclopramide, a medicine associated with tardive dyskinesia after prolonged cumulative exposure.[10] The authors used prescription claims data to estimate duration of therapy and the extent of therapy beyond the maximum time period of 12 weeks evaluated in the clinical trials and recommended in the label. During the study period, almost 80% of approximately 200 000 participants had only one episode of therapy. The length of the longest episode for most patients (85%) varied from 1 to 90 days, yet 15% of the patients appeared to have received prescriptions for metoclopramide for a period longer than 90 days. Cumulative therapy for longer than 90 days was recorded for almost 20% of the patients. These data indicate that a substantial percentage of patients were taking metoclopramide for longer than the recommended duration of treatment. The manufacturer was subsequently required to add additional warning to the product's label, cautioning against prolonged use.[11]

The case of acetaminophen (paracetamol) illustrates use of public-health databases compiled not specifically for pharmacoepidemiologic purposes, but which can offer important insights into how medicines are used. Acetaminophen is one of the most widely used medications, and is available in several single-ingredient and multi-ingredient, over-the-counter, and prescription products. Although acetaminophen is generally safe when used as directed, misuse and overdose can cause acute liver failure, sometimes resulting in transplantation or death. In fact, acetaminophen is the leading cause of drug-induced liver failure in the United States.[12] Overdoses can be either intentional or unintentional. For the regulator, understanding the conditions that lead to toxicity is of critical importance, since interventions can only be designed if the conditions of use are well understood. To understand the patterns of use that give rise to toxicity, Nourjah and colleagues at the FDA examined several national databases to quantify this problem.[13] 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 overdoses 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 annually to unintentional overdose. These databases provide quantitative data on both the overall magnitude of acetaminophen overdoses in the United States as well as on the proportion of the overdoses that occur unintentionally. Data such as these are critical for the regulator. By providing a context for the environment in which toxicity occurs, these data allow for the design of targeted risk mitigation interventions, as well as for

monitoring their impact once interventions have been implemented. In the UK, regulators have undertaken reduction of package size as a risk-mitigation measure.[14] In the US, the FDA has asked drug manufacturers to limit the strength of acetaminophen in prescription acetaminophen-containing products to 325 mg per tablet, capsule, or other dosage unit.[15] The FDA is also considering what steps may be taken to minimize the occurrence of serious liver damage related to over-the-counter acetaminophen use.

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 sometimes not enough simply to add a warning to a product label.[16] The impact of this regulatory action must be assessed. Because many regulatory actions recommend certain conditions of use for a medicine, it is possible to measure adherence to these conditions, rather than directly measuring the health outcome of interest. As formal risk management programs become increasingly used to manage specific serious risks of medicines, scientifically rigorous assessments of these programs will be needed to insure that the goals of the program are being met (see Chapter 29). Pharmacoepidemiology is critical to these endeavors, as it can relate drug usage both to patient characteristics and patterns of use of other medicines as well as to patient outcomes. The case below illustrates the measurement of adherence to labeled recommendations regarding contraindicated medicines.

Cisapride, an agent used to promote motility of the upper gastrointestinal tract, was first marketed in the United States in September 1993. Between that time and April 1996, the US FDA received reports of 34 patients who developed torsade de pointes while on cisapride, and 23 reports of patients developing a prolonged QT interval while on cisapride. Further investigation revealed that these cases were associated with elevated serum cisapride levels, which in many cases were due to the concomitant administration of CYP 3A4 inhibitors.[17] In 1995, the FDA required that Boxed Warning be added that contraindicated the use of cisapride in patients taking CYP 3A4 inhibitors. A Dear Health Care Professional Letter was also sent at that time. As more information gathered that concomitant use of medicines that prolong the QT interval could be dangerous in patients taking cisapride, the FDA in 1998 strengthened the Boxed Warning to include contraindication of medicines that prolong the QT interval. During the time between the first Boxed Warning and the second, cisapride use continued to grow. A second Dear Health Care Professional Letter was sent in June 1998 to 800 000 health-care professionals. To assess the impact of these regulatory actions on cisapride use, Smalley and colleagues used electronic data from two managed care programs and one state Medicaid program to compare the proportion of cisapride users for whom cisapride use was contraindicated, based on the product's label, in two time periods, the 1 year before the June 1998 action (July 1997–June 1998) and the 1 year afterward (July 1998–June 1999).[18] In the databases examined the proportions of cisapride users for whom the medication was contraindicated in the year before the June 1998 action were 26%, 30%, and 60%. In the year following the actions, the corresponding proportions were 24%, 28%, and 58%. The authors concluded that FDA's actions had no substantial effect on the use of cisapride in patients for whom it was contraindicated. Weatherby and colleagues used outpatient pharmacy claims data from a large New England (USA) insurer to address the same question, and found a 58% decline in the rate of co-prescription of medications that were explicitly mentioned as contraindicated medicines in the Dear Health Care Professional Letter.[19] However, there was no similar decline in the co-prescription of medicines that were mentioned as examples of member of drug class or implied as drug class members. In 2000, the manufacturer, in consultation with FDA, decided to discontinue marketing the product, and made the product available to patients who met clearly defined eligibility criteria through an investigational limited access program.

Integrating data from multiple sources

Central to the role of a drug regulator is determining the benefit–risk balance of a medicine and, to date, pharmacoepidemiology has been particularly central to determining the risk part of this balance. The entirety of the pharmacoepidemiologic armamentarium is involved. Case reports, case series, non-randomized epidemiologic studies, clinical trials, and meta-analyses are amongst the most common techniques used. These topics are covered extensively in this book (see Chapters 3, 5, 23, 33, 36, 37, and 40), and their technical aspects will not be discussed here. While the regulator must be familiar with these techniques, the regulator's approach to pharmacoepidemiology must be one that integrates findings from various data sources that have been analyzed with various techniques. Other sources of data, such as clinical pharmacology findings (see Chapter 2) and the results of animal toxicology studies, may also contribute to the overall body of data. Indeed, the ability to integrate findings from diverse data sources depends, in large part, on a thorough understanding of the data and the methods of analysis. Beyond the technical issues, the regulator is faced with determining the significance of the data at hand, and what regulatory action, if any, must be taken.

There is no one single approach to synthesizing data from multiple data sources. Rather, a careful and structured approach must be taken in each case. Considerations include the risk being studied, the magnitude of the effect seen, the sources of the data used, the control for bias and confounding in each study, the robustness of each finding, biological plausibility, and prior findings. There is no simple algorithm that can be used in all situations.

Standard hierarchies of evidence have been published, though these are not always relevant for the drug regulator in making a decision about the safety of a medicine. For example, case reports and case series, which are usually accorded the lowest status in an evidence hierarchy (see Chapter 3), may be the only practical way to determine that rare, serious adverse events are associated with a medicine. Thus, strict reliance on clinical trials to determine that a certain medicine is associated with aplastic anemia or acute liver failure would be misguided. In other situations, a safety outcome might be sufficiently uncommon, or occur with such long latency, that it might not be feasible to study it in clinical trials. If the outcome may nonetheless be seen in patients with the condition for which the medicine is given, case reports will not be helpful. In these situations, carefully designed observational studies would be appropriate.

Recently, there has been renewed attention to the role of synthesizing evidence from many sources, and to the notion that the traditional hierarchies of evidence may not always be appropriate for assessment of drug safety.[20] Because of their rigid, step-wise approach to evidence, hierarchies are not well suited to integrating diverse sources of information. To make the issue more challenging, data from experimental settings (i.e., clinical trials) may yield one estimate of effect, while observational data may yield a quantitatively different estimate. These differences in effect size between study settings may be sufficient to result in different interpretations of the risk–benefit balance.[21] In practice, however, data on a safety issue do come from diverse sources. The challenge for regulators, who must use the data to make public health decisions, is to integrate the available information optimally. This may be particularly challenging when trying to integrate and balance clinical trial results relating to benefit and pharmacoepidemiologic study results relating to risk. The best ways to do this have not yet been established, and will likely be a topic of much interest in the coming decade.

Building capacity and collaboration in pharmacoepidemiology

Pharmacoepidemiology is a complex field, and relies on epidemiology, clinical pharmacology, pharmacy, medicine, statistics, and other disciplines, for its full execution (see also Chapters 2 and 3). 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 health-care data, which experts in the above fields may not have. As more and more drug safety questions arise that require expertise in pharmacoepidemiologic and appropriate data, it is crucial that there be sufficient capacity, in the form of well-trained pharmacoepidemiologists, and that there also be appropriate venues for collaboration. Regulators can play a role in reaching these goals.

To strengthen the postapproval monitoring of medicines, the European Medicines Agency (EMA) developed The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP)[22] (see also Chapter 6). 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 Database of Research Resources, launched in early 2010, is a publically available searchable, electronic index of European research resources in pharmacovigilance and pharmacoepidemiology. The database covers both centers and networks, as well as data sources in the European Union. The ENCePP Code of Conduct, also released in 2010, provides a set of rules and principles for pharmacovigilance and pharmacoepidemiologic studies, with regard to best practices and transparency. A parallel Checklist of Methodological Standards for ENCePP Study Protocols allows researchers to be aware of and consider important methodologic considerations. The ENCePP e-Register of Studies, also released in 2010, provides an important transparency, tracking, and results dissemination tool focused on pharmacoepidemiologic studies.[23] The ENCePP project illustrates one way that a regulatory agency can be involved in building capacity.

The US FDA, as one of its commitments under the re-authorization 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 epidemiologic best practices and provides guidance on carrying out scientifically sound observational studies using quality data resources."[24] This task illustrates another mechanism through which regulatory agencies can promote the field of pharmacoepidemiology.

In addition to providing guidance on best practices for pharmacoepidemiologic studies, FDA has funded formal pharmacoepidemiologic studies. One early effort in this area was funding for the development and use of the Computerized On-Line Medicaid Pharmaceutical Analysis and Surveillance System (COMPASS), a computerized database of inpatient and outpatient medical claims and outpatient pharmacy claims of participants in the Medicaid program, a health benefit program in the US for certain low-income individuals and families who meet defined eligibility criteria.[25] As additional large, population-based databases have become available for pharmacoepidemiologic studies, FDA's funding has focused on experts in pharmacoepidemiology who have access to relevant data and who collaborate with FDA epidemiologists on drug safety questions of mutual interest, and on funding studies using many of the data sources described elsewhere in this book.

Because pharmacoepidemiology depends on many areas of expertise, fostering collaboration is another potential role that regulators can play. Through a project funded by the Innovative Medicines Initiative (IMI) in March 2010, the EMA, along with the national drug regulatory agencies from Denmark, Spain, and the United Kingdom, have 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.[26] Topic areas covered by PROTECT include enhancing data collection from consumers; improving early and proactive signal detection from spontaneous reports, electronic health records, and clinical trials; developing, testing, and dissemination 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.

In the US, the Agency for Healthcare Research and Quality, one of FDA's sister agencies in the Department of Health and Human Services, consults with FDA on funding and running the Centers for Education and Research on Therapeutics (CERTs), a national initiative to conduct research and provide education that advances the optimal use of therapeutics (see also Chapter 6).[27] The program consists of 14 academic research centers, a Coordinating Center, a Steering Committee, and partnerships with public and private organizations. Each of the 14 centers focuses its research efforts on a particular population or therapeutic area.

FDA's Sentinel Initiative (see Chapter 30) also represents an example of a program sponsored by a regulatory agency that seeks to advance pharmacoepidemiology though a collaborative effort.[28] The goal of the Sentinel Initiative is to create a sustainable, linked system of electronic health-care databases to investigate safety questions about FDA-regulated medical products. The use of health-care data in this way raises many questions of public interest, including questions of governance, privacy, data standards, and public disclosure of results. In view of these issues, FDA has sought extensive stakeholder input as it works with outside organizations to develop Sentinel. In addition to the logistical issues above, the fundamental premise of Sentinel—that data from many sources can be used to address a drug safety question—implies that a collaborative effort is needed for the success of this project.

Conclusion

In conclusion, pharmacoepidemiology is a critical discipline in the activities of a drug regulatory agency. Key issues for regulators at this time include optimally using pharmacoepidemiology across the lifecycle of a product; developing clear, robust, and transparent methods of integrating data from multiple sources to arrive at sound, evidence-based conclusions; and promoting the building of capacity in the field of pharmacoepidemiology. These efforts are interdependent and depend not only on the efforts of regulatory agencies, but also on collaborations with academia and industry.

References

1. 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.
2. European Medicines Agency. *Orphan Designation*. 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 June 2011).
3. Schieppati A, Henter JI, Daina E, Aperia A. Why rare diseases are an important medical and social issue. *Lancet* 2008; **371**: 2039–41.
4. 21CFR314. Code of Federal Regulations Title 21, 314.510. *Approval based on a surrogate endpoint or on an effect on a clinical endpoint other than survival or irreversible morbidity*.
5. Commission Regulation (EC) No 507/2006 of 29 March 2006 on the conditional marketing authorisation for medicinal products for human: http://ec.europa.eu/health/documents/eudralex/vol-1/index_en.htm (accessed June 2011).
6. McKee AE, Farrell AT, Pazdur R, Woodcock J. The role of the US Food and Drug Administration review process: Clinical trials endpoints in oncology. *Oncologist* 2010; **15** (Suppl 1): 13–18.
7. US Department of Health and Human Services. Food and Drug Administration. *Guidance for Industry: Diabetes Mellitus—Evaluating Cardiovascular Risk in New Anitdiabetic Therapies to Treat Type 2 Diabetes*, December 2008.
8. *Volume 9A - Pharmacovigilance for Medicinal Products for Human Use*, Part I.3. http://ec.europa.eu/health/documents/eudralex/vol-9/index_en.htm (accessed June 2011).
9. McMahon AW, Pratt RG, Hammad TA, Kozlowski S, Zhou E, Lu S, *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**: 921–33.
10. Kaplan S, Staffa JA, Dal Pan GJ. Duration of therapy with metoclopramide: a prescription claims data study. *Pharmacoepidemiol Drug Saf* 2007; **16**: 878–81.

11. FDA. FDA Requires Boxed Warning and Risk Mitigation Strategy for Metoclopramide-Containing Drugs Agency warns against chronic use of these products to treat gastrointestinal disorders, Feb. 26, 2009. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2009/ucm149533.htm (accessed June 2011).
12. Bower WA, Johns M, Margolis HS, Williams IT, Bell BP. Population-based surveillance for acute liver failure. *Am J Gastroenterol* 2007; **102**: 2459–63.
13. 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.
14. Hawton K, Townsend E, Deeks J, Appleby L, Gunnell D, Bennewith O, *et al*. Effects of legislation restricting pack sizes of paracetamol and salicylate on self poising in the United Kingdom: before and after study. *BMJ* 2001; **322**: 107.
15. Department of Health and Human Services. Food and Drug Administration. Prescription drug products containing acetaminophen; actions to reduce liver injury from unintentional overuse. *Fed Regist* 2011; **76**: 2691–7.
16. US Department of Health and Human Services. Food and Drug Administration. *Guidance for Industry: Development and Use of Risk Minimization Action Plans*, March, 2005.
17. Wysowski DK, Bacsanyi J. Cisapride and fatal arrhythmia. *N Engl J Med* 1996; **335**: 290–1.
18. Smalley W, Shatin D, Wysowski DK, Gurwitz J, Andrade SE, Goodman M, *et al*. Contraindicated Use of Cisapride: Impact of Food and Drug Administration Regulatory Action. *JAMA* 2000; **284**: 3036–9.
19. Weatherby LB, Nordstrom BL, Fife D, Walker AM. The impact of wording in "Dear doctor" letters and in black box labels. *Clin Pharmacol Ther* 2002; **72**: 735–42.
20. Rawlins M. *De Testimonio*: on the evidence for decisions about the use of therapeutic interventions. *Lancet* 2008; **372**: 2152–61.
21. van Staa TP, Smeeth L, Persson I, Parkinson J, Leufkens HG. Evaluating drug toxicity signals: is a hierarchical classification of evidence useful or harmful? *Pharmacoepidemiol Drug Saf* 2008; **17**: 475–84.
22. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. www.encepp.eu (accessed June 2011).
23. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. *Electronic Register of Studies*. http://www.encepp.eu/encepp/studiesDatabase.jsp (accessed June 2011).
24. Department of Health and Human Services. US Food and Drug Administration. *PDUFA IV Reauthorization Performance Goals and Procedures; Fiscal Years 2008 through 2012*. http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm119243.htm (accessed June 2011).
25. Strom BL, Carson JL, Halpern AC, Schinnar R, Snyder ES, Stolley PD, *et al*. Using a claims database to investigate drug-induced Stevens–Johnson syndrome. *Stat Med* 1991; **10**: 565–76.
26. Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium. http://www.imi-protect.eu/index.html (accessed June 2011).
27. Agency for Healthcare Research and Quality. *CERTs Overview*. http://certs.hhs.gov/about/certsovr.htm (accessed June 2011).
28. US Department of Health and Human Services. US Food and Drug Administration. *The Sentinel Initiative—National Strategy for Monitoring Medical Product Safety*, May 2008. http://www.fda.gov/downloads/Safety/FDAsSentinelInitiative/UCM124701.pdf (accessed June 2011).

CHAPTER 9
Pharmacoepidemiology and the Law

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

Introduction

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 of these cases, the existence of governing law helps influence the way people act. For example, knowing of the existence of tort liability rules should incentivize people to take care to prevent accidents from happening.

Pharmacoepidemiologists in their daily work 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 these cases, pharmacoepidemiologic studies may provide the scientific underpinning for the claim, helping determine whether a manufacturer is at fault. 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 (or "intellectual property"), often through the use of patents.

This chapter outlines three of the most important 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 pharmacoepidemiologists. In each example, US law is used as the paradigm (given the legal background of the author) but, since much of the discussion is based on principles that are generally similar in other highly developed 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 reported cases have included the non-steroidal 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

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

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 itself. 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 happened to purchase the preparation directly from the manufacturer. The emergence of product liability law altered that state of affairs. Product liability law as a legal theory generally arose in the US around the time of the Industrial Revolution as a variation of tort law, permitting consumers harmed by the many products sold widely in interstate commerce to seek redress for their injuries.[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 make out 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, courts found that 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. Courts 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 of whether the defendants followed customary practices or exercised reasonable precautions is moot. The strict liability principle permitted plaintiffs to argue that they should receive 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, 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 of which 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 a warning defect—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 (Second) of Torts, which included an important annotation relevant to prescription drugs. In Comment k, 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] The Restatement effectively excluded most prescription drugs from the 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 courts implemented the Comment k principle and offered pharmaceutical manufacturers a blanket protection from strict liability for manufacturer or design defect claims.[7] However, 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 Declomycin, a tetracycline-based 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 was published with the intention to clarify the question about liability for design defects. It re-emphasized 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 FDA-approved drugs 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 which also largely hinges on the appropriateness of the accompanying warnings. Practically speaking, the ultimate disposition of a case filed under 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.

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

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.[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 that 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 of a fact. Actual knowledge can be demonstrated by a showing that the manufacturer was cognizant of reasonable information suggesting a particular risk that it did not pass on to consumers, for example, where a defendant possesses data about certain adverse events that were not disclosed. In the case of selective serotonin reuptake inhibitors (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 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 1993–1996; it showed no benefit of the drug over placebo and five cases (out of 93) of suicidal ideation, as compared to one case out of 89 in the placebo arm and one 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 receiving antidepressants for possible worsening of depression or suicidality" 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 by that time. 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.

Warning

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 as high as >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 acknowledge this clinically important difference until more than a year after the drug was marketed.[19]

Second, 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 many years before this article was published but delayed warning about them and sought to limit their dissemination.[21] Thus, a primary question in lawsuits arising from 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 prokinetic agent linked to potentially fatal cardiac side effects. It was reported that the manufacturer and the FDA negotiated for 5 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]

Third, warnings must be of appropriately urgent tone. In the case of rofecoxib (Vioxx®), a COX2-protective 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 another NSAID in a population of patients with rheumatoid arthritis (but no known cardiovascular disease).[23] However, in the published manuscript, the authors presented the results as suggesting that the difference was due to a substantial cardioprotective effect of naproxen, rather than a harmful effect of rofecoxib. 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 label.[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 progestational agent (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

Another major issue in a pharmaceutical product liability case is whether the product at issue 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 study populations to test those hypotheses on the basis of studies using probabilistic thinking (i.e., a *p* value). However, legal causation usually requires a clear causal chain from event to outcome, in an individual. The legal standard for causation is therefore challenged by product liability cases, where probabilistic evidence links drugs to injuries.[26] Courts struggle with the question of legal causation in these cases on two distinct levels: general and specific causation.

General causation addresses whether a product is capable of causing a particular injury in the population of patients like the plaintiff. The basic 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 rule 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 appropriately 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 build the bridge between an association and general causation (see Table 9.1).

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

To demonstrate *specific causation*, a plaintiff must show that the product in question caused the alleged injury in the individual 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, even after proving general causation. 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, "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 that resulted 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.

Learned intermediary defense

If a plaintiff demonstrates a prima facie case of product liability, manufacturers have a few defenses that they can offer. 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 the best position to understand and relay complex information about possible side effects. Physicians are also uniquely 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 the prescribing physician.[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 that physicians can sufficiently grasp, 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 some 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 manufacturers lack means of interacting with patients other than through physicians. 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 will usually play the role of the expert witness in product liability cases. Pharmacoepidemiologists can help judges and juries understand data about drugs and help determine whether risk information was acted upon appropriately. Experts are usually called on to

describe the current state of knowledge about the adverse event at issue, and may be asked to perform 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 markers 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 a more basic 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 opinions, lack a reliable method, or fail to apply the methods to the facts.[45] Notably, there is considerable skepticism about the effectiveness of courts as a gatekeeper for expert witnesses, with commentators citing the argument that judges lack 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 all prescription drugs are relatively safe and show efficacy for their intended indication before the drug can be sold on the US market.[47] At the time of approval, the FDA also endorses an official drug label, which presents a description of the basis of the drug's efficacy as well as safety concerns that have emerged during the period of preapproval testing.[48] The label, which is generally written by the manufacturer and approved by the FDA, has legal significance as well. For example, because the FDA restricts manufacturer communication about unapproved (or "off-label") drug uses, the label determines what a pharmaceutical manufacturer communicates to physicians and the public about its product.[49] The FDA requires the manufacturer to mention important warnings that are in the official label when marketing its product, but does not require manufacturers to mention warnings that are not in the label. At the time of drug approval, the label represents the FDA's best judgment about risks that warrant disclosure and how to describe those risks.

For most of its history, the FDA has regulated the drugs sold in the US without any significant role in product liability litigation brought by consumers injured by FDA-approved drugs.[50] The agency's non-interventionist 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 immediately started experiencing agitation, confusion, and suicidal thinking, ultimately leading him to take his own life 1 week later.[51] The plaintiffs claimed that the manufacturer failed to warn appropriately about

the risks of suicidal behaviors. 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 Constitution prevents states from imposing legal requirements (in this case, via a tort action in state court) in direct contradiction to federal law.[52] Under new political leadership, 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 FDA argued that an adverse court ruling forced companies to add warnings not approved by the FDA could upset the delicate risk–benefit balancing that went into the construction of the drug label and could lead to over-warning and ultimately under-use of an effective drug.

The major deficiency in the arguments in favor of FDA preemption is that they 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 may not have a complete picture of the safety of drugs, even at the time the label 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 drug label warnings. After the withdrawal of rofecoxib, Sandra Kweder, Deputy Director of the FDA's Office of New Drugs, said in testimony in 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 FDA's hand somewhat, ensuring compliance can still involve a lengthy and resource-intensive legal process.

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 promote the effectiveness of their drugs and 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. Manufacturers faced with this conflict of interest can make poor decisions that adversely affect public health. For example, when safety issues do emerge, 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 label 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 the argument gained some

judicial traction, as 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 label, "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 FDA preemption claim. 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 in 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 manufacturer's knowledge of the drug's risk 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 first that it was impossible to comply with the federally approved label and, second, 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 six–three 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.

Pharmacoepidemiology and contract law

Many studies in the field of pharmacoepidemiology emerge from collaborations between individuals at different institutions. Two researchers may bring different types of expertise to a project or different resources.[67,68] For example, researchers may have all the computing power they need, but require access to a certain external database in order to address a particular question. Collaborations may occur among academic centers, between non-profit 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 non-profit 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; a recent 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.[69] In countless cases, contract research in pharmacoepidemiology has led to important public-health findings and changes in health-care 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.[70,71] These results can be explained by choices made in trial design, where subjective decisions about comparators[72] or the inclusion or exclusion of certain variables or potential confounders in epidemiologic and economic studies can affect the ultimate results of the trial.[73] 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."[74] 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.[75]

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.[76] Company documents suggest that the clinical trial database was exclusively controlled by the company, which provided limited access to the investigators.[77] 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 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 that they delineate who controls access to the data.

Finally, there have been numerous conflicts over so-called "gag clauses" that prevent contract investigators from publishing their ultimate results.[78] 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,[79] 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."[80] In the recent case of the cholesterol-lowering drug ezetimibe (Zetia®), the outside investigator leading a large-scale clinical trial found 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.[81]

Such problems are not unique to private industry contracts. In the US, a recent 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.[82]

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.[83] For example, a median of 1% of research

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.

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.[84]

A few expert bodies have offered recommendations on legal guidelines for the conduct of contract research.[85] The best known and most authoritative has 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.[86] 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). Problematic language includes overly broad confidentiality clauses, clauses that define and assign ownership of intellectual property, 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 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."[87]

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 14th and 15th 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."[88] 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.[89] 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 non-obvious. These criteria aim to ensure that patents cannot be awarded for inventions that already exist, or small improvements on those inventions that are obvious to a person of ordinary skill in the field. 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 formalized 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 rationale for pursuing patents. First, patents provide an incentive for investment in research and 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.[90] 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 his invention available to the public and fully describes it, so that others can use it and potentially improve on it in creating subsequent innovation. However, patents can also be controversial. Patents over scientific research tools have been implicated in barriers to effective cooperation,[91] enhanced secrecy among researchers,[92] and restrictions on availability of the products of research to patients.[93]

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,"[94] or purely mental process.[95] On the other hand, 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]

In the past few years, patents related to the practice of pharmacoepidemiology have been obtained by individuals,[96] start-up companies,[97] and even large health-care data collectors such as Microsoft.[98] For example, one patent was recently awarded 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. [99]

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 has 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.[100] In another case, a patent-seeker in the field has argued that researchers should patent the adverse reactions discovered in pharmacoepidemiologic studies in order to enhance funding from for-profit pharmaceutical companies that might be interested in novel and non-obvious processes that link drugs and adverse events.[101] 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."[102]

Recently, the US Supreme Court has stepped into the controversy over process patents. In 2008, the Court of Appeals for the Federal Circuit, the highest 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.[103] 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."[104] 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.[105]

Despite the Supreme Court's reluctance to draw a bright-line separating patentable from non-patentable processes, there may still be limits to attempts to patent certain epidemiology-related methods.[106] For example, the Federal Circuit Court has recently reviewed a set of patents related to methods of immunization and adverse effect detection.[107] In that case, an inventor had secured a patent on a method of using adverse event data regarding vaccine administration to inform subsequent health-care 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.[108]

The Federal Circuit found this claim too abstract to be patentable because it was "unrelieved by any movement from principle to application."[109] 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.

Conclusion

Legal issues intersect with the practice of pharmacoepidemiology in a number of 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 a drug label completely precludes providing additional warnings about adverse effects where the warnings are warranted by the data. 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, pharmacoepidemiologists may decide to try to patent their research methods, but should weigh the risks and benefits of this form of intellectual property.

References

1. Kesselheim AS, Avorn J. The role of litigation in defining drug risks. *J Am Med Assoc* 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, Torts (4th ed.) § 99 (1971).
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*. West Group, 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. http://www.oag.state.ny.us/press/2004/aug/aug26a_04_attach1.pdf (accessed June 2011).
13. Kondro W, Sibbald B. Drug company experts advised staff to withhold data about SSRI use in children. *CMAJ* 2004; **170**: 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. *Worsening Depression and Suicidality in Patients Being Treated With Antidepressants*, 22 March 2004. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHeathcareProfessionals/PublicHealthAdvisories/ucm161696.htm (accessed June 2011).
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**. *N 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. *N Engl J Med* 2007; **356**: 2457–71.
21. Committee on Finance, United States Senate, Max Baucus, Chairman, Charles Grassley, Ranking Member. *Staff Report on GlaxoSmithKline and the Diabetes Drug Avandia*, 2010.
22. Harris G, Koli E. Lucrative drug, danger signals and the FDA. *NY Times*, 10 Jun, 2005.
23. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, *et al*. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med* 2000; **343**: 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, 2000. http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/$file/sciman00.pdf (accessed June 2011).
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**: 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. *N 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**: 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**: 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, Jan 2009. GAO Report, GAO-09-271.
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 June 2011).
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 Stated 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. Kong SX, Wertheimer AI. Outcomes research: collaboration among academic researchers, managed care organizations, and pharmaceutical manufacturers. *Am J Manag Care* 1998; **4**: 28–34.
68. 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**: 129–35.
69. Zinner DE, Campbell EG. Life-science research within US academic medical centers. *JAMA* 2009; **302**: 969–76.
70. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Int Med* 1986; **1**: 155–8.
71. 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.
72. 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.
73. 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.
74. Bodenheimer T. Uneasy alliance: clinical investigators and the pharmaceutical industry. *N Engl J Med* 2000; **342**: 1539–44.
75. Gottlieb S. Firm tried to block report on failure of AIDS vaccine. *BMJ* 2000; **321**: 1173.
76. Home PD, Pocock SJ, Beck-Nielsen H, Curtis PS, Gomis R, Hanefeld M, *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.
77. Nissen SE. Setting the RECORD straight. *JAMA* 2010; **303**: 1194–5.
78. Steinbrook R. Gag clauses in clinical-trial agreements. *N Engl J Med* 2005; **352**: 2160–2.
79. Olivieri NF, Brittenham GM, McLaren CE, Templeton DM, Cameron RG, McClelland RA, *et al*. Long-term safety and effectiveness of iron-chelation therapy with deferiprone for thalassemia major. *N Engl J Med* 1998; **339**: 417–23.
80. Shuchman M. Legal issues surrounding privately funded research cause furore in Toronto. *CMAJ* 1998; **159**: 983–6.
81. Berenson A. Accusations of delays in releasing drug results. *NY Times*, 1 Apr, 2008.
82. Association of American Universities/ Council on Government Relations. Restrictions on research awards: troublesome clauses 2007/2008. Jun 2008. http://www.aau.edu/policy/scientific_openness.aspx?id=6872 (accessed June 2011).
83. Schulman KA, Seils DM, Timbie JW, Sugarman J, Dame LA, Weinfurt KP, *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.
84. Mello MM, Clarridge BR, Studdert DM. Academic medical centers' standards for clinical-trial agreements with industry. *N Engl J Med* 2005; **352**: 2202–10.
85. Halperin SA, Scheifele D, Duval B, Law B, Ward B, Bjornson G, *et al*. Canadian Association for Immunization Research and Evaluation (CAIRE) guidelines for industry-sponsored clinical trial and epidemiologic contract research. *Hum Vaccin* 2005; **1**: 140–2.
86. International Committee of Medical Journal Editors. *Uniform Requirements for Manuscripts Submitted to Biomedical Journals; Writing and Editing*, Oct 2008. http://www.icmje.org/urm_full.pdf (accessed June 2011).
87. Bodenheimer T. Uneasy alliance: clinical investigators and the pharmaceutical industry. *N Engl J Med* 2000; **342**: 1539–44.
88. U.S. Const. art. I, § 8, cl. 8.
89. 35 U.S.C. § 101.
90. Porcelli FP, Dragseth JA. *Patents: a Historical Perspective*. Minneapolis, MN: Fish and Richardson, 2002.
91. Heller MA, Eisenberg RS. Can patents deter innovation: the anticommons in biomedical research. *Science* 1998; **280**: 698–701.

92. Blumenthal D, Campbell EG, Gokhale M, Yucel R, Clarridge B, Hilgartner S, *et al*. Data withholding in genetics and the other life sciences: prevalences and predictors. *Acad Med* 2006; **81**: 137–45.
93. Merz JF, Kriss AG, Leonard DGB, Cho MK. Diagnostic testing fails the test. *Nature* 2002; **415**: 577–9.
94. *Diamond v. Diehr*, 450 U.S. 175 (1981).
95. *Gottshalk v. Benson*, 409 U.S. 63 (1972).
96. US Patent 7,461,006.
97. US Patent 7,653,639.
98. US Patent 7,406,453.
99. US Patent 7,650,262.
100. Mullin J. The fight of his life. *IP Law and Business*, Apr–May 2009: 44–8.
101. Classen JB. Enhanced funding of pharmacoepidemiology through patenting the disclosure of adverse event information. *Pharmacoepidemiol Drug Saf* 2006; **15**: 390–3.
102. Walker AM. More lawyers, more bureaucrats, less information on drug safety. *Pharmacoepidemiol Drug Saf* 2006; **15**: 394–5.
103. In re Bilski, 545 F.3d 943 (Fed. Cir. 2008) (en banc).
104. In re Bilski, 545 F.3d 943 (Fed. Cir. 2008) (en banc).
105. *Bilski v. Kappos*, 130 S. Ct. 3218 (2010).
106. Kesselheim AS. The Supreme Court, process patents, and medical innovation. *N Engl J Med* 2009; **361**: 2303–6.
107. *Classen Immunotherapies v. Biogen Idec* (Fed. Cir., 2008).
108. US Patent 5,723,283.
109. *Classen Immunotherapies v. Biogen Idec* (Fed. Cir. 2011).

PART III

Sources of Data for Pharmacoepidemiologic Studies

CHAPTER 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

Introduction

Just about 50 years ago a concerned obstetrician from Australia sent a letter to the editor of the *Lancet* medical journal describing severe malformations in babies born to several of his patients, and posing the question: "Have any of your readers seen similar abnormalities in babies delivered of women who have taken [thalidomide] during pregnancy?"[1] An accompanying editor's note gave the news that the drug's manufacturer (Distillers) planned to withdraw the product from marketing based on "reports from two overseas sources possibly associating thalidomide ('Distaval') with harmful effects on the fetus." Sadly, the number of babies born with thalidomide-related deformities has been estimated to exceed 10 000 worldwide[2] before the scope and consequences of this adverse drug reaction were adequately appreciated and effectively prevented. Today, perhaps in part as a result of this tragic occurrence, drug safety surveillance and regulatory decision-makers are more vigilant in detecting safety signals and more active in alerting prescribers about serious adverse drug effects, largely through established pharmacovigilance reporting systems worldwide.

In recent years, the term "pharmacovigilance" has become widely used to denote postmarketing 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 drug-related problems."[3]

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 lifecycle, 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

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, Fifth Edition. Edited by Brian L. Strom, Stephen E. Kimmel, Sean Hennessy.

(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 Food and Drug Administration (FDA) reviews the professional labeling (package insert), to ensure that the product's uses and risks are explained adequately.

Though 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 the elderly, 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 of 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

Central to an understanding of postmarketing pharmacovigilance reporting systems are 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,[4] defines an adverse event as follows:

> 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.[5]

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. . . ."[6]

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.

Overview of pharmacovigilance reporting systems

The goal of a postmarketing, or postapproval, safety program is to identify drug-related adverse events (AEs), or adverse drug reactions (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 the 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 into 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 postmarket 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 using the medicine. If this sign or symptom is not listed in the product's approved labeling, patients and health-care 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 Food and Drug Administration in the United States or other national or regional authorities, as appropriate.

In the postmarketing period, the investigation of AEs/ADRs is a multidisciplinary one. 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 that drug from marketing based on "a markedly increased reporting rate of fatal rhabdomyolysis" compared to the other drugs in the statin class.[7] Additional confirmation of the unacceptably high risk of rhabdomyolysis with

cerivastatin was eventually available 3 years later when results of a well-designed epidemiologic study were published.[8] 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.[9]

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 the information to regulators. National reporting systems are typically run by regulatory agencies (e.g., the US FDA runs the MedWatch program[10]) 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 to 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 to 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 the 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 (see also below, in the section National Pharmacovigilance Systems).

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 health-care records) may provide some opportunity to monitor the safety of a marketed medicine (see Section IIIB and Chapter 30 of this book), 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 health-care professionals, patients, and consumers want to make a 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 health-care professionals in postmarketing safety assessment. Though the practices and systems for

health-care professionals to report AEs/ADRs vary from region to region, the quality of reports is always dependent on the details provided by health-care professionals.

Because most AE/ADR reporting systems rely on health-care professionals, patients, and consumers to submit reports voluntarily, it is generally recognized that there is substantial under-reporting of AEs/ADRs via current reporting systems. Two survey-based studies conducted in the US in the 1980s, one in Maryland[11] and the other in Rhode Island,[12] examined physician reporting to FDA, and concluded that fewer than 10% of AEs/ADRs were reported to FDA. These studies were conducted prior to the development of the current MedWatch program[13] 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 actually 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 FDA of cases of hospitalized rhabdomyolysis associated with statin use was estimated[14] using a projected estimate of the number of such cases in the United States and comparing it to the number of reports of statin-associated hospitalized rhabdomyolysis in the FDA's Adverse Event Reporting System (AERS), 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 incidence rates to national estimates of statin use. Across four statins (atorvastatin, cerivastatin, pravastatin, and simvastatin), the estimated overall extent of adverse event 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-adverse 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.[15] The International Conference on Harmonization (ICH) E2B standard 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. 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: (i) patient characteristics, (ii) reaction(s) or event(s), (iii) results of tests and procedures relevant to the investigation of the patient, (iv) drug(s) information, and (v) a narrative case summary and further information.[16]

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.[17]

Characteristics of a good quality case report have been published.[16,17] 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 not 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 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 a recently published analysis of AERS data because the reports were "not interpretable."[18] Similarly, in a recent published analysis of the clinical spectrum and causality assessment of spontaneous reports of telithromycin-associated hepatotoxicity, several potentially relevant cases were not included in the case series because of "insufficient clinical or laboratory information to enable assessment."[19]

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 the 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 de-challenge (the resolution of the AE/ADR when the medication is withdrawn) and re-challenge (the re-development of the AE/ADR when the drug is re-introduced), if available, can be invaluable.

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 resource-poor 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 are 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 resource-poor countries there is currently not enough capacity for effective safety monitoring, and the improved access to new medicines adds additional strain on already overburdened or non-existent pharmacovigilance systems. In a recent survey of pharmacovigilance systems in low- and middle-income countries, 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.[20] 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 World Health Organization (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 will be focused on sustainable development, covering not only the implementation of reporting systems, technical support, and training of health-care professionals, but also improvements in governance and infrastructure to support pharmacovigilance activities.

The perceived responsibility of health-care 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,[21] a symbol used to denote a drug 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 New Zealand, the Intensive Medicines Monitoring Programme monitors cohorts of all patients taking selected new drugs, and specifically requests that all clinical events be reported, not just suspected adverse drug reactions.[22] 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 region, other high income countries, and 33 of 55 low- and middle-income countries responding to a 2008 survey,[20] pharmaceutical companies that hold marketing authorizations are obligated to report adverse events or adverse drug reactions 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[10] 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 non-serious, 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 (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 region, 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 post-marketing 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, which are connected to major university hospitals. In the United Kingdom, there are four regional centers connected to university hospitals, which have a 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 State Food and Drug Administration, SFDA. In India, an improved pharmacovigilance system is being developed by the Central Drugs Standard Control Organization,

under the Ministry of Health and Family Welfare. In the first year, up to 40 medical institutes are expected to participate in pharmacovigilance activities, with further increases in a phased manner until, in 2013, all medical colleges will be linked to four to six regional offices.

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 that include suspected AE/ADR reports derived from a postmarketing reporting system, as well as suspected AE/ADR reports from other sources, such as the published medical literature, and sometimes certain types of serious adverse events (SAEs; those considered related to 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 SWEDIS database (Sweden), the Sentinel database (United Kingdom), the Adverse Event Reporting system (AERS) 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[23] in the European Union (run by the European Medicines Agency, EMA) and VigiBase[24] pooling data from the approximately 100 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 28 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 the WHO Drug Dictionary Enhanced.[25] 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).[26] 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, biologicals, 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 "non-serious" 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."[27] 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 if 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. There are a variety of specific algorithms that have been applied to safety signal detection in AE/ADR databases.[28,29] A full discussion of the statistical principles underlying these methods is beyond the scope of this chapter.

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, 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.[30] 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 know as the Gamma Poisson Shrinker or the Multi-item Gamma Poisson Shrinker).[31]

Review of case reports

The review of individual case reports of AEs/ADRs is a complex process that has been described elsewhere[32,33] (also see Chapter 33), and has been the subject of public discussion.[34] 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 the medicine, 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 in order to determine if 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.[35–40] For example, if the AE/ADR of interest is one that 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's disease is much higher in persons over age 60 years than it is in persons below that age.[41] In this situation, review of a report describing a myocardial infarction in a 70-year-old patient on an anti-parkinsonian 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 a AE/ADR is most straightforward when the suspected AE/ADR: (i) is rare in the population when the medication is not used, (ii) is not a manifestation of the underlying disease, (iii) has a strong temporal association with drug administration, and (iv) 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 the past four decades.[35,37] In general, the approach relies on a systematic review of each case report to ascertain the temporal relationship between drug intake and the development of the adverse reaction, an assessment of any co-existing diseases or medications that could confound the relationship between the medicine and the AE/ADR, the clinical course after withdrawing the drug (de-challenge), and the clinical course after re-introduction of the drug (re-challenge), 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.[36,42] The World Health Organization has developed a qualitative scale for categorizing causality assessments.[3]

In the development of a case series, once the individual cases are reviewed, it is important to integrate the findings across the cases in an effort 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 is it 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 was 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 many AEs/ADRs do 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 made with caution, since such comparisons are highly sensitive to variation in AE/ADR reporting and thus it is necessary to take into account 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 are 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.[43–46] 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 health-care professionals to report suspected AEs/ADRs to national pharmacovigilance centers, drug regulatory authorities, and/or manufacturers. Such systems allow for direct engagement of health-care 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 de-challenge and re-challenge. 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 such reporting systems. In the United States, where consumers and patients can report either to the manufacturer or directly to the FDA, the percentage of reports in 2009 that originated from consumers was 46%. While consumer and patient-generated reports might not have the same level of medical detail as those provided by health professionals, subsequent follow up with health professionals may be possible in potentially important cases, so that more complete clinical information can be obtained.

Scope

The scope of a postmarketing safety reporting system is quite broad. The system can cover all medicines used in the populations, 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 lifecycle. Thus, AEs/ADRs recognized late in a product's lifecycle, such as those resulting from prolonged exposure to a medicine, can, in theory, be ascertained. In practice, such ascertainment is difficult to achieve, because health-care 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 health-care 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 the development of the AE/ADR. The quality of the report, as described earlier in this chapter, is critical for an informative and meaningful review of the individual case report. 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 the report 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 of AEs/ADRs 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.

Non-uniform 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 the drug–event 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 has been thought to peak in the second year after approval and declines thereafter, even though the drug may be used more widely. This phenomenon, known as the Weber effect, was originally described in relation to non-steroidal anti-inflammatory medicines.[47] A recent analysis[48] 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. Specifically, this analysis did not confirm that the number of reports increased toward the end of the second year and declined thereafter. Rather, the analysis 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 year's period, it was found that the adjusted number of reports was highest in the first years after approval and declined thereafter. Thus, the frequency of AE/ADR reports per estimated unit of drug utilization is not likely to be constant over time.

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 there was 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 make 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 "de-duplicated".

Particular applications

Felbamate and aplastic anemia

The case of aplastic anemia associated with felbamate therapy illustrates the role that case reports can play in the assessment of a previously unknown AE/ADR during the postapproval period. Felbamate is an anticonvulsant agent approved for use in the United States on July 29, 1993. Preapproval studies showed no evidence of significant, non-reversible hematologic abnormalities.[49] Within about 1 year of approval, 20 cases of aplastic anemia, three of them fatal, had been reported in the United States. Review of the case reports suggested a causal role for felbamate. An estimated 100 000 patients had taken felbamate during this time.[50] While the true incidence of aplastic anemia in patients taking felbamate cannot be calculated because case ascertainment is likely incomplete, the minimum rate is 20/100 000/year, or 200/million/year. By contrast, the population background rate of aplastic anemia

is low, about 2/million/year[43] Thus, the observed cases of aplastic anemia suggest that aplastic anemia is at least 100 times more frequent in patients taking felbamate than in the general population. Based on this finding, the FDA and the manufacturer recommended that patients not be treated with felbamate unless the benefits of the drug were judged to outweigh the risk of aplastic anemia. A subsequent review of 31 case reports of aplastic anemia in patients taking felbamate[43] using the criteria of the International Agranulocytosis and Aplastic Anemia Study (IAAAS), established that felbamate was the only plausible cause in three cases, and the most likely cause in 11 cases. For the remaining nine cases, there was at least one other plausible cause. The authors concluded that the "most probable" incidence of aplastic anemia was estimated to be to 127 per million. Because aplastic anemia is uncommon in the population and because it is generally the result of a medication or other toxin, a careful analysis of a case series can establish the relationship of a drug to aplastic anemia.[43,49,50]

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.[51] 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 drugs' manufacturers. The results of the analysis were presented at a subsequent Pediatric Advisory Committee meeting in March 2006[52] and were also later published in a peer-reviewed journal.[53] 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 that were studied.[54]

Serious medical errors, a focus of increasing concern[55,56] (see also Chapter 45), can also be identified with targeted surveillance of spontaneous reporting systems. Steps to prevent or mitigate the likelihood of such errors are often feasible. For instance, after FDA approval of Kapidex (dexlansoprazole) in January 2009, reports of dispensing errors due to confusion with the drugs Casodex (bicalutamide, indicated for advanced prostate cancer) and Kadian (morphine sulfate, an opioid analgesic) were received. FDA was especially concerned by reports of the inadvertent administration of Casodex to women, as Casodex is contraindicated in women and is classified as pregnancy category X. In order to avoid future medication errors, the proprietary name of dexlansoprazole delayed-release capsules (Kapidex) was changed to Dexilant in March 2010.[57]

Data mining signals

Below are two examples of WHO Program signals identified by data mining applied to the WHO Global Individual Case Safety Report Database, VigiBase. The disproportionality measure used by the UMC is the Information Component (IC), originally introduced through the Bayesian Confidence Propagation Neural Network (BCPNN), which is a logarithmic measure of the disproportionality between the observed and expected reporting of a drug–ADR 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.

Topiramate and glaucoma

Topiramate was approved in the US in 1996 as an anticonvulsant drug.[58] 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 this 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. (Note: the Signal document is a UMC publication 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). 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. FDA issued a warning in the revised labeling October 1, 2001.[58]

Abacavir and myocardial infarction

Abacavir is a nucleoside analog reverse transcriptase inhibitor (NRTI) used to treat HIV and AIDS. It was approved by the US FDA in 1998. In the second quarter of 2000, an association between abacavir and myocardial infarction was identified as a potential signal using data mining of VigiBase. Starting from 11 cases in June 2000, the number of cases had increased to 34 by the second quarter of 2004, although seven were identified as possible duplicates. The signal was included in the Signal document in May 2005, with the comment that "with these data the association between abacavir and myocardial infarction should not be ruled out, and further studied." In April 2008, published results of the large international D:A:D (Data Collection on Adverse Events of Anti-HIV Drugs) cohort study[59] following 33 347 patients, described a relative rate (95% CI) for myocardial infarction associated with recent exposure to abacavir of 1.90 (1.47–2.45); however, a definite causal link could not be proven. In July 2008, FDA reported in an Early Communication about an ongoing safety review of abacavir and didanosine, initiated by the findings from the D:A:D study. Also in July 2008, abacavir labeling was revised to include a warning of the potential increased risk for myocardial infarction. It was pointed out that no excess risk for myocardial infarction was observed in a sponsor-conducted pooled analysis of abacavir clinical trials (n = 9600). However, although available data from the observational cohort and from clinical trials "remain inconclusive", the FDA advised that the underlying risk for coronary heart disease be considered when prescribing antiretroviral therapies and precautions taken to minimize all modifiable risk factors.[60–62]

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. The first, from Nigeria, was presented by Adeline Osakwe at a meeting in Rabat, Morocco, 2009; the second, from Sri Lanka, in Geneva, 2005.

Safety concerns on adulterated excipients in manufacture of pharmaceutical products in Nigeria

In November 2008, the Nigeria National Pharmacovigilance Center received reports of pediatric patients with unexplained acute renal failure (ARF), resulting in 84 deaths. An investigation was made into the cause of the illness, and effective risk mitigation and risk minimization measures were put in place by the regulatory authority to prevent further injuries.

The investigation included hospital surveillance (retrospective analysis as well as increasing physician awareness to encourage reporting of new cases), standardized case-family interviews (collection of demographic, clinical, and drug-exposure data), pertinent sample collection, and laboratory analyses. The results of the investigation revealed 111 cases of ARF and 84 (75.7%) fatalities suspected to be related to paracetamol-containing medications that were contaminated with diethylene glycol

(DEG). Twenty-eight (32%) of the deaths were administered My Pikin®, a brand of paracetamol/diphenhydramine mixture used for teething problems in infants.

Immediate interventions included issuance of public alerts, removal of the product from the supply chain, closure of the manufacturing plant, and distribution of fomepizole antidote, while long-term measures include stricter control of excipients/chemicals and strengthening of enforcement of cGMP for manufacturer of pediatric formulations.

Outbreak of health-care-associated meningitis

A representative from Sri Lanka highlighted how the reporting of suspected AEs/ADRs may contribute significantly to patient safety when information is analyzed and followed up and root causes identified.

There were 14 cases with three deaths reported as fungal contaminations of anesthetic drugs/syringes. The date of onset ranged between the weeks of June 27 and July 25, 2005. Case-patients were identified following exposures in three different health-care facilities. Case-patients did not share a common health-care facility, personnel, or surgery session.

Microbiology indicated *Aspergillus fumigatus* as the infectious agent in three of seven case-patients. There was evidence of the organism in patients who received spinal anesthesia in two of the three hospitals involved. Two medications were common to six of the cases, but they were not contaminated. A large variety of injection devices used in the health care facilities and coming from three manufacturers were contaminated with *Aspergillus fumigatus*. Substandard storage conditions may have constituted the mode of contamination of these devices. There had been a lack of storage facilities because of very large donations in response to a tsunami. Two of the three syringes implicated were tsunami donations.

The future

Spontaneous AE/ADR reporting is an important component of drug safety surveillance. The widespread availability of electronic health-care 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 provide routinely a clinically meaningful narrative that explains concisely the clinical course of the AE/ADR and its relationship to medication usage.

Postmarketing safety reporting systems depend on the involvement of health-care 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 active surveillance and the use of large health-care 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. McBride WG. Thalidomide and congenital abnormalities. *Lancet* 1961; **2**(721): 1358.
2. McBride W. Specialist life—William McBride. *Eur J Obstet Gynecol Reprod Biol* 2001; **95**: 139–40.
3. Uppsala Monitoring Centre. *Pharmacovigilance. Definitions*. http://who-umc.org/DynPage.aspx?id=97224&mn1=7347&mn2=7252&mn3=7257 (accessed June 2011).
4. ICH harmonized tripartite guideline. Post-approval safety data management: definitions and standards for expedited reporting. *Lancet* 2003; **336**: 156–8.

5. *ICH Harmonized Tripartite Guideline Clinical Safety Data Management: Definitions and Standards for Expedited Reporting E2A*. Current Step 4 version, 27 October 1994. http://www.ich.org/home.html (accessed June 2011).
6. US Food and Drug Administration. 21 CFR 314.80. 4-1-2009. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=314.80 (accessed June 2011).
7. Staffa J, Chang J, Green L. Cerivastatin and reports of fatal rhabdomyolysis. *N Engl J Med* 2002; **346**: 539–40.
8. Graham DJ, Staffa JA, Shatin D, Andrade SE, Schech SD, La Grenade L, *et al*. Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs. *JAMA* 2004; **292**: 2585–90.
9. Psaty B, Furberg C, Ray W, Weiss N. 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.
10. US Food and Drug Administration. *MedWatch: The FDA Safety Information and Adverse Events Reporting System*, 2010. http://www.fda.gov/Safety/MedWatch/default.htm (accessed June 2011).
11. Rogers AS, Israel E, Smith CR, Levine D, McBean AM, Valente C, *et al*. Physician knowledge, attitudes, and behavior related to reporting adverse drug events. *Arch Intern Med* 1988; **148**: 1596–600.
12. Scott HD, Rosenbaum SE, Waters WJ, Colt AM, Andrews LG, Juergens JP, *et al*. Rhode Island physicians' recognition and reporting of adverse drug reactions. *R I Med J* 1987; **70**: 311–6.
13. Kessler D. Introducing MEDWatch: A new approach to reporting medication and device adverse effects and product problems. *JAMA* 1993; **269**: 2765–8.
14. McAdams M, Staffa J, Dal Pan G. Estimating the extent of reporting to FDA: a case study of statin-associated rhabdomyolysis. *Pharmacoepidemiol Drug Saf* 2008; **17**: 229–39.
15. US Food and Drug Administration. *Guidance for Industry—E2B(M): Data Elements for Transmission of Individual Case Safety Reports*, 2005. http://www.fda.gov/RegulatoryInformation/Guidances/ucm129428.htm (accessed June 2011).
16. US Food and Drug Administration. *The ICH Guideline on Clinical Safety Data Management: Data Elements for Transmission of Individual Case Safety Reports*, 5-12-2009. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm129399.pdf (accessed June 2011).
17. US Food and Drug Administration. *Guidance for Industry: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment*, 2005. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126834.pdf (accessed June 2011).
18. McMahon A, Pratt R, Hammad T, Kozlowski S, Zhou E, Lu S, *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.
19. Brinker AD, Wassel RT, Lyndly J, Serrano J, Avigan M, Lee WM, *et al*. Telithromycin-associated hepatotoxicity: Clinical spectrum and causality assessment of 42 cases. *Hepatology* 2009; **49**: 250–7.
20. Olsson S, Pal S, Stergachis A, Couper M. Pharmacovigilance activities in 55 low- and middle-income countries: a questionnaire-based analysis. *Drug Saf* 2010; **33**: 689–703.
21. MHRA. *Black Triangle Scheme*, 2010. http://www.mhra.gov.uk/index.htm (accessed June 2011).
22. Coulter D. The New Zealand Intensive Medicines Monitoring Programme. *Pharmacoepidemiol Drug Saf* 1998; **7**: 79–90.
23. Eudravigilance. *Eudravigilance. Pharmacovigilance in EEA*, 2010. http://eudravigilance.ema.europa.eu/highres.htm (accessed June 2011).
24. Uppsala Monitoring Centre. *Vigibase™ Services*, 2010. http://www.umc-products.com/DynPage.aspx?id=4910&mn=1107 (accessed June 2011).
25. Uppsala Monitoring Centre. *WHO Drug Dictionary Enhanced*, 2010.
26. MedDRA- Medical Dictionary for Regulatory Activities and MSSO—Maintenance and Support Services Organization, 2010. http://www.meddramsso.com/ (accessed June 2011).
27. World Health Organization. *The Importance of Pharmacovigilance. Safety Monitoring of Medicinal Products*. Geneva, 2002.
28. Bate A, Lindquist M, Edwards I, Orre R. A data mining approach for signal detection and analysis. *Drug Saf* 2002; **25**: 393–7.
29. Szarfman A, Machado S, O'Neill R. 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**: 381–92.
30. Sachs B, Riegel S, Seebeck J, Beier R, Schichler D, Barger A, *et al*. Fluoroquinolone-associated anaphylaxis in spontaneous adverse drug reaction reports in Germany: differences in reporting rates between

individual fluoroquinolones and occurrence after first-ever use. *Drug Saf* 2006; **29**: 1087–100.
31. Almenoff J, Tonning JM, Gould AL, Szarfman A, Hauben M, Ouellet-Hellstrom R, *et al*. Perspectives on the use of data mining in pharmaco-vigilance. *Drug Saf* 2005; **28**: 981–1007.
32. Institute of Medicine. Committee on the Assessment of the US Drug Safety System. Chapter 2: Natural history of a drug. In: *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. National Academies Press, 2006.
33. US Food and Drug Administration. The FDA's drug review process: Ensuring drugs are safe and effective. *FDA Consumer Magazine*, July–August 2002. Pub No. FDA05-3242.
34. FDA. *Drug Safety and Risk Management Advisory Committee Meeting*, May 18–19, 2005. http://www.fda.gov/OHRMS/DOCKETS/98fr/05-7458.htm (accessed June 2011).
35. Jones J. Causality assessment of suspected adverse drug reactions: a transatlantic view. *Pharmacoepidemiol Drug Saf* 1992; **1**: 251–60.
36. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA *et al*. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; **30**: 239–45.
37. Meyboom RHB, Hekster YA, Egberts ACG, Gribnau FWJ, Edwards IR. Causal or casual? The role of causality assessment in pharmacovigilance. *Drug Saf* 1997; **17**: 374–89.
38. Benichou C, Danan G. Experts' opinion in causality assessment: results of consensus meetings. *Drug Inf J* 1991; **25**: 251–5.
39. 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.
40. Miremont G, Haramburu F, Begaud B, Péré J, Dangoumau J. Adverse drug reactions: physicians' opinions versus a causality assessment method. *Eur J Clin Pharmacol* 1994; **46**: 285–9.
41. Bower J, Maraganore D, McDonnell S, Rocca W. Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976–1990. *Neurology* 1999; **52**: 1214.
42. Naranjo C, Lanctôt K. Microcomputer-assisted bayesian differential diagnosis of severe adverse reactions to new drugs: a 4-year experience. *Drug Inf J* 1991; **25**: 243–50.
43. Kaufman D, Kelly J, Anderson T, Harmon D, Shapiro S. Evalution of case reports of aplastic anemia among patients treated with felbamate. *Epilepsia* 1997; **38**: 1265–9.
44. Wysowski D, Corken A, Gallo-Torres H, Talarico L, Rodriguez E. Postmarketing reports of QT prolongation and ventricular arrythmia in association with cisapride and Food and Drug Administration regulatory actions. *Am J Gastroenterol* 2001; **96**: 1698–701.
45. Wysowski D, Bacsanyi J. Cisapride and fatal arrhythmia. *N Engl J Med* 1996; **335**: 290–1.
46. Graham D, Green L, Senior J, Nourjah P. Troglitazone-induced liver failure: a case study. *Am J Med* 2003; **114**: 299–306.
47. Weber J. Epidemiology of adverse reactions to nonsteroidal anti-inflammatory drugs. In: Rainsford K, Velo G, eds. *Advances in Inflammation Research*. New York: Raven Press, 1984, pp. 1–7.
48. McAdams M, Governale L, Swartz L, Hammad T, Dal Pan G. 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.
49. Pennell P, Ogaily M, Macdonald R. Aplastic anemia in a patient receiving felbamate for complex partial seizures. *Neurology* 1995; **45**: 456–60.
50. Nightingale S. Recommendation to immediately withdraw patients from treatment with felbamate. *JAMA* 1994; **272**: 995.
51. US Food and Drug Administration. *Pediatric Advisory Committee*, June 30, 2005. www.fda.gov/ohrms/dockets/ac/05/briefing/2005-4152b2.htm (accessed June 2011).
52. US Food and Drug Administration. *Pediatric Advisory Committee*, March 22, 2006. www.fda.gov/ohrms/dockets/ac/06/briefing/2006-4210B-Index.htm (accessed June 2011).
53. Mosholder A, Gelperin K, Hammad T, 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.
54. FDA MedWatch. *Attention Deficit Hyperactivity Disorder (ADHD) Drug Products: Medication Guides Required to Alert Patients to Possible Cardiovascular and Psychiatric Risks*, February 21, 2007. www.fda.gov/medwatch/safety/2007/safety07.htm#ADHD (accessed June 2011).
55. Kohn L, Corrigan J, Donaldson M. *To Err is Human: Building a Safer Health System*. Washington, DC: Institute of Medicine, 1999.
56. Honig P, Phillips J, Woodcock J. How many deaths are due to medical errors? *JAMA* 2000; **284**: 2187–8.

57. Hamilton-Stokes D. FDA approves brand-name change for dexlansoprazole. *Drug Topics*, May 15, 2010.
58. Thambi L, Kapcala LP, Chambers W, Nourjah P, Beitz J, Chen M, *et al*. Topiramate-associated secondary angle-closure glaucoma: a case series. *Arch Ophthalmol* 2002; **120**: 1108.
59. D:A:D Study Group; Sabin CA, Worm SW, Weber R, Reiss P, El-Sadr W, Dabis F, *et al*. Use of nucleoside reverse transcriptase inhibitors and risk of myocardial infarction in HIV-infected patients enrolled in the D:A:D study: a multi-cohort collaboration. *Lancet* 2008; **371**: 1417–26.
60. Jong E, Meijers J, van Gorp E, Spek C, Mulder J. Markers of inflammation and coagulation indicate a prothrombiotic state in HIV-infected patients with long-term use of antiretroviral therapy with or without abacavir. *AIDS Res Ther* 2010; **7**: 9.
61. Lang S, Mary-Krause M, Cotte L, Gilquin J, Partisani M, Simon A, *et al*.; Clinical Epidemiology Group of the French Hospital Database on HIV. Impact of individual antiretroviral drugs on the risk of myocardial infarction in human immunodeficiency virus-infected patients: a case-control study nested within the French Hospital Database on HIV ANRS cohort CO4. *Arch Intern Med* 2010; **170**: 1228–38.
62. Costagliola D, Lang S, Mary-Krause M, Boccara F. Abacavir and cardiovascular risk: reviewing the evidence. *Current HIV/AIDS Reports* 2010; **7**: 127–33.

CHAPTER 11
Overview of Automated Databases in Pharmacoepidemiology

Brian L. Strom
Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

Once hypotheses 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 have the ability to detect drug effects with an incidence as low as 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. A study of this size would be 95% certain of observing at least one case of any adverse effect that occurs with an incidence of 3 per 10 000 or greater (see Chapter 4). However, studies this large are expensive and difficult to perform. Yet, these 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 three decades have seen a growing use of computerized databases containing medical care data, so 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, 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 patient's unique identifier, that the records are updated on a regular basis, and that the records are verifiable and are reliable. The ability to conduct medical chart review to confirm outcomes is also a necessity

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

for most studies, 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 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. In the current chapter, we will introduce these resources, presenting some of the general principles that apply to them all. In Chapters 12 to 18 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 existed and 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, in Europe, medical record databases have been developed for use by researchers, and similar databases have been developed in the US more recently.

Claims and other administrative databases

Claims data (Chapters 13, 14, and 17) arise from a person's use of the health-care system (see Figure 11.1). 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.

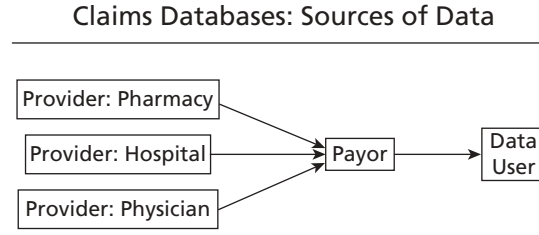


Figure 11.1 Sources of claims data.

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 (see Chapter 14). Indeed, there have also been numerous validity checks on the drug data in claims files that showed that the drug data are of extremely high quality, that is 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 41).

The quality of disease data in these databases is somewhat less perfect. If a patient is admitted to a hospital, the hospital charges for the care and justifies that charge by assigning International Classification of Diseases–Ninth Revision–Clinical Modification (ICD-9-CM) codes and a Diagnosis Related Group (DRG). The ICD-9-CM 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 reason to provide incorrect ICD-9-CM codes. In fact, most hospitals have mapped each set of ICD-9-CM 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 does not usually depend on the actual diagnosis, but rather on the procedures administered during the outpatient medical encounter, and these procedure codes indicate the intensity of the services provided. Thus, there is no incentive for the practitioner to provide incorrect ICD-9-CM 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, for example data from US Health Maintenance Organizations (Chapter 12) or other data sources (Chapter 18). The characteristics of these data are similar in many ways to those of claims data.

Medical record databases

In contrast, medical 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 itself. In many instances, this is replacing the paper medical 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 General Practice Research Database (GPRD), described in Chapter 15, along with the newer database, The Health Improvement Network (THIN). As general practice databases, these contain primarily outpatient data. In addition, recently there are new inpatient electronic medical record databases available (Chapter 16).

Medical record databases have unique advantages. Importantly among them is that the validity of the diagnosis data in these databases is better than that in claims databases, as these data are being used for medical care. When performing a pharmacoepidemiologic study using these databases, there is no need to validate 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. These include 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, that is for claims databases, information is available on all medical care provided, regardless of who the provider was. As indicated above, this can be a problem though for medical records databases. 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 external other electronic databases (e.g., death records, maternal–child records, police accident records), to expand the capabilities and scope of

research. This requires using 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 medical record 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 smoking, alcohol, date of menopause, etc., all of which can be of great importance to selected research questions. This argues that one either needs access to patients or access to 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.

A major other 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 dis-enrollment of plan members. Another source of instability of the population is when patients transfer out of the system due to death or moving away. 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 1 or 2 years after the patient's last recorded entry in the database). Alternatively, strategies can be adopted for selecting stable populations within a particular database, for example, by examining patterns of prescription refills for chronically used medications. Of course, the largest such data system, that is 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 health care which is provided (see Chapter 14).

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 not of importance.

Some results from studies that utilize these databases may not be generalizable, for example on health-care utilization. This is especially relevant for databases created by data from a population that is atypical in some way, for example US Medicaid data (see Chapter 14).

Finally, 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 health-care providers. Thus, providers' electronic health records are inherently incomplete, and need to be linked to administrative data in order to be useful for quality research.

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;
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 defined by the ICD-9-CM coding system, such as Stevens–Johnson syndrome;
4 descriptive studies, if the population studied is skewed;
5 delayed drug effects, wherein patients can lose eligibility in the interim;
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. Inasmuch 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 advent in the US of FDA's Sentinel system, destined to exceed 100 million individuals (Chapter 30). 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 12–18, we will review the details of a number of these databases. The databases selected for detailed 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 (see Chapter 18) 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.

CHAPTER 12
Health Maintenance Organizations/ Health Plans

Susan E. Andrade,[1] Marsha A. Raebel,[2] Denise Boudreau,[3] Robert L. Davis,[4] Katherine Haffenreffer,[5] Pamala A. Pawloski,[6] Sengwee Toh,[5] and Richard Platt[5]

[1] Meyers Primary Care Institute, University Of Massachusetts Medical School, Worcester, MA, USA

[2] Institute For Health Research, Kaiser Permanente Colorado, Denver, CO, USA

[3] Group Health Research Institute, Seattle, WA, USA

[4] Center For Health Research—Southeast, Kaiser Permanente Georgia, Atlanta, GA, USA

[5] Harvard Pilgrim Health Care Institute and Department of Population Medicine, Harvard Medical School, Boston, MA, USA

[6] HealthPartners Research Foundation, Bloomington, MN, USA

Introduction

Health maintenance organizations (HMOs), often also referred to as health plans, are an important resource for certain types of pharmacoepidemiologic research. We use the term HMO here to refer to health-care delivery systems that assume responsibility for preventive and therapeutic health services to a defined population, in contrast to insurance systems which simply pay for the care provided by others. Their salient features for research typically include most of the following: (i) responsibility for the care for a defined population, (ii) information about diagnoses and procedures resulting from care delivered in both ambulatory and inpatient settings, access to full inpatient and outpatient medical records, and outpatient pharmacy dispensing data, and (iii) the ability to interact with providers and members/patients. HMOs with these attributes are well positioned to make extensive use of routinely collected electronic data, including both administrative claims and electronic health record data, to perform an array of record linkage studies, such as cohort, nested case–control, and case-based studies. Researchers are also able to supplement this information with full-text record review, and to obtain additional information from providers or patients.

In this chapter, we describe the interplay of health-care systems, operational principles, capabilities, expertise, and data resources that support research across multiple therapeutics areas. These same characteristics also make these entities good environments in which to conduct an array of public-health surveillance activities, epidemiologic research, health-services research, and clinical trials.

Description

Data resources

Administrative and clinical data sets maintained by HMOs and used for clinical care, payment, and operational purposes, serve as major resources for many epidemiologic studies. The principal sources of information are administrative and claims data, and electronic medical records. The categories of data are described first; additional information specific to electronic medical records follows. Descriptions of data elements commonly found are as follows.

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

Administrative and claims data

Demographic data and membership status

Date of birth and gender are routinely available. HMOs also maintain detailed information on dates of enrollment and termination of membership, residence, and change in benefit plans. This membership data allows characterization of well-defined populations; that is this information can be used to qualify and identify subjects with incident drug use in inception cohorts and to follow-up and censor individuals should they leave the health plan (and study observation). Linkage of these data to census data (geocoding) can provide proxy measures of race/ethnicity and socioeconomic status.

Outpatient drug exposure

HMOs often offer pharmacy benefits to members, providing a strong financial incentive (except possibly for low-cost prescriptions) for members to receive their drugs through a mechanism that results in a claim for reimbursement or a dispensing record. For health plans with integrated pharmacies, the pharmacy dispensing information is obtained directly from the dispensing record. Each drug claim or dispensing record contains a unique National Drug Code (NDC) that identifies the active ingredient(s), dose, and formulation, as well as the route of administration, amount dispensed, days of supply, and prescriber. Information about copayment amounts is also typically available. Drugs that are administered intravenously in special clinics or office visits can be identified by either the dispensing record or a special designation, such as the Health Care Financing Agency Common Procedure Coding System (HCPCS) codes, for the visit during which the drug was administered. However, ascertainment of drug exposure using these automated dispensing records may be incomplete. Information on drugs that are used during hospitalizations is usually not available at most HMOs. Over-the-counter medication use is generally not captured. It is unclear how completely HMO records capture the dispensing of low-cost prescriptions. At many HMOs, prescribing, a measure of clinicians' intent, can be identified through orders in electronic medical records, described below.

Diagnoses

Diagnoses associated with hospitalizations or ambulatory visits can be identified from automated claims or health plans' electronic ambulatory medical records. Most diagnoses are recorded in International Classification of Diseases, 9th revision Clinical Modification (ICD-9-CM) codes. Plan-specific ambulatory diagnosis codes are sometimes used, and must be translated into ICD-9-CM codes for research purposes.

Procedures/special examinations

Hospital and ambulatory procedures (laboratory tests, radiology examinations, endoscopy examinations, surgeries, or others) are coded according to the ICD-9-CM, Current Procedural Terminology (CPT), HCPCS, or plan-specific systems.

Electronic medical records (EMRs)

Electronic medical records (EMRs) contain information that can substantially enhance pharmacoepidemiologic studies. EMRs based in HMOs are especially valuable since their information can be linked to the administrative and claims data described above. Outside the HMO setting, EMR data are more difficult to use, since they are typically restricted to care provided in a location or locations that share a practice's (or hospital's or network's) EMR. For systems that are not based in HMOs or do not have the capability to link to claims data, there is usually no way to identify eligible person-time, to obtain information about dispensing of medications provided outside the practice, or to know about care from other providers, even when the EMR is network-based.

The EMRs use a controlled medical terminology (e.g., based upon ICD-9-CM and CPT systems) to document patient assessments and procedures. Further, they support clinician order entry for pharmacy, laboratory, radiology, referrals, and provider-to-provider messaging. Finally, clinical notes, diagnoses, orders, and test results are archived, allowing their use for research and other analyses. Additionally, data including race, ethnicity, and health behaviors (e.g. smoking status, body

mass index, and alcohol use) increasingly are available.

HMO-based investigators are also developing and testing natural language processing (NLP) methods to extract valuable, research-quality data from clinical text and demonstrating the great potential of these technologies in applied settings.[1–6]

Full-text records

HMOs typically have direct access to traditional paper outpatient and inpatient medical records for their enrolled members if needed for years prior to when their EMRs began.

Additional HMO research databases and registries

Health-plan-based research centers have developed a variety of databases that support research, including many condition-specific disease registries (see Chapter 21). These additional databases and registries can leverage administrative, claims, and EMR data. Such embedded registries can provide efficient approaches for studying questions related to the natural history of disease conditions, the effectiveness or adverse effects of medications, and treatment patterns.

Examples of such registries include tumor registries, which can obtain their data from case identification systems within the local sites and site-owned hospitals (routine review of all pathology and cytology reports, oncology consultations, etc.) and/or linkage to data from state registries or Surveillance, Epidemiology and End Results (SEER) information. At some sites, data are collected in a standardized format using nationally standardized definitions and formats set by the North American Association of Central Cancer Registries.

Basic diabetes registries (e.g. diagnosis, laboratory, medication data) or enhanced diabetes registries (e.g. adherence information, patient-reported outcomes) are also available in a number of HMOs, some extending over a decade. Disease registries are discussed further in Chapter 21.

In addition, cardiovascular disease research databases or registries are maintained at a number of HMOS, including databases of patients diagnosed with acute myocardial infarction, individuals with stroke, patients with heart failure and preserved left ventricular systolic function, patients with diagnosed and undiagnosed hypertension, atrial fibrillation, and venous thromboembolism management and long-term outcomes.[7–10] The data include information on co-morbidities, medications, therapeutic interventions, laboratory testing, complications, and other health-care utilization. Registries of specific cardiac drug therapies and devices and procedures have also been developed, including implantable cardiac defibrillators, cardiac catheterizations, cardiac surgeries, and warfarin-treated patients with atrial fibrillation or venous thromboembolism.[11] Other examples of research databases maintained by HMOs to augment routinely available data include ones for asthma, immunizations, HIV/AIDS, hepatitis B and C, joint replacement, renal replacement/chronic kidney failure, bariatric surgery, and robotic surgery.

Data development procedures for multi-HMO projects

The work invested to date in developing and using HMOs as venues for pharmacoepidemiologic investigation has been substantial. It has been particularly challenging to conduct studies that require populations larger than a single health plan's. Thus, considerable effort has been devoted to developing standards that allow more efficient use of single organization's data for repeated use, and also facilitate the implementation of a single research protocol in multiple health plans. We describe here a general approach to coordination of HMO-based pharmacoepidemiologic investigation.

For multicenter research activities, a coordinating center typically works with the lead investigator to prepare a data development plan. Computer analytic programs, for instance written using SAS software (SAS Institute, Inc. www.sas.com), are developed at the coordinating center or lead project site and executed at each site to identify study subjects and generate study-specific data elements. Standard processes and code have been developed to identify drug and define exposures, clinical outcomes, potential confounders, prior medical

history, and medical procedures of interest. For studies evaluating medication exposures, specific drugs can be identified using National Drug Code (NDC) or Generic Product Identifier (GPI) codes as a method of standardizing drug-related data collection across sites. Reusable programs can to calculate person-time at risk, identify time of event occurrence, and to calculate co-morbidity indices such as the Charlson index or the Chronic Disease Score.[12–14] The central development of analytic programs ensures consistent implementation of the protocols across multiple health plans and decreases the programming costs and data variability at each site.

For descriptive studies in which the reportable results are aggregate data, site-specific tables can be generated by standard programs from source data at each site and then combined to support the final analyses; there is no need in such studies to transfer person-level data. When there is a compelling need to generate a combined data set with person-level data for complex statistical analyses, it is then necessary to share only the minimum data necessary. The following processes can be used to minimize the amount of detailed information that leaves the originating health plan:

- Age: date of birth is not shared. Age as of an index event, e.g. hospital discharge with a diagnosis of acute myocardial infarction, is calculated in study-appropriate groups, e.g. 5-year intervals for adults.
- Dates: occurrences of all other events of interest are specified relative to the index event, such as the dispensing of a beta-blocker 35 days after a hospitalization for acute myocardial infarction, or the dispensing of an antiepileptic drug 150 days before delivery of a baby. While preserving the temporal sequence of events of interest, data prepared in this fashion have no dates at the person-level that are shared outside of the health plan.
- Specific diagnosis and procedure codes are grouped into clinically meaningful entities.
- Drugs of interest are grouped into therapeutic classes using the American Hospital Formulary Service (AHFS) or other standardized classification systems, or individual drugs (generic names) using the NDC system.
- Composite scores, such as the Chronic Disease Score,[14] the Charlson co-morbidity index,[12,13] or propensity scores[15,16] are computed at the sites, and the scores (or, if necessary, their components) are transmitted. The individual data elements, such as specific drug dispensings that contribute to the score, are not shared.

A randomly generated study identifier (ID) replaces the unique health plan identifier for each study subject. The crosswalks between the HMO identifier and the study ID are securely stored at each site. The study IDs are not reused, so that the same person identified in different studies would not carry the same study ID. In some cases, for example, if the unit of analysis is a specific clinical event, more than one study ID is used for a person in a single study. In this example, an analysis of patterns of laboratory testing associated with drug prescribing might assign different study IDs to the same person for each of two target drugs.

Strengths

The potential for large and diverse defined populations, the varied delivery models and practice patterns, together with automated claims and EMR data, access in many plans to full-text medical records, access to providers, and ability to work with the health plans' members are valuable assets for research requiring large, diverse populations and delivery systems. Large cohorts can be identified to evaluate incidence of rare events and to study these events among sufficiently large numbers of patients with certain co-morbidities, such as hypertension, diabetes, and heart failure. Health plans' access to medical records for confirmation of clinical events is essential for certain studies. The increasing richness of the computerized clinical data offer an advantage over many other data sources. Lastly, the research centers' ability to contact their health-plan enrollees for participation in studies is extremely valuable for clinical research. This ability to contact members facilitates enrollment of patients in clinical trials (see Chapter 36) and pharmacogenomics research projects (see Chapter 34), and also enables the

conduct of studies which include patient interviews or questionnaires to provide information on patient behaviors (e.g. physical activity, over-the-counter medication use), beliefs, and knowledge not captured in the administrative and clinical HMO databases.

Coordination and data development infrastructure enables both observational and interventional studies to be conducted efficiently across health plans, using a distributed data processing model for such studies. Creation of institutional review board (IRB) reliance agreements among health plans further reduces the barriers to coordination between HMOs by allowing delegation of human subjects' committee review and approval to the lead project site, simplifying administrative requirements and human subjects' oversight.

Weaknesses

The most important limitation of HMO data sources is perhaps the absence of population groups that are uninsured. These missing groups are typically highly enriched for individuals who do not qualify for insurance because they are unemployed. A consequence of this is that HMO populations can be less diverse than the population as a whole, with a smaller proportion of socioeconomically disadvantaged individuals.

Some health plans have a smaller fraction of the elderly than the general population, because individuals who are 65 years and older disproportionately receive their care through Medicare fee-for-service programs (see Chapter 14).

Because membership is often associated with employment status, turnover of the population can occur when employers contract with different health plans or when individuals change jobs.

Some benefits, such as mental health services, may be "carved-out," that is contracted en bloc to another organization, and thus not be captured in detail by HMOs' record systems.

Some health plan benefit plans cap the amount of certain services, such as physical therapy, and would thus not necessarily capture services that the individual pays for individually.

Some individuals may have more than one source of health insurance, for instance if spouses each have separate family coverage. In this case, an HMO may not capture all care an eligible individual receives.

While the data are rich in elements related to health care, information on race and ethnicity, indicators of socioeconomic status, and lifestyle factors (e.g. smoking and alcohol consumption) are not yet readily available for all members. However, data on some of these factors, such as smoking status, are increasingly available in EMRs. While the completeness of many data elements available in EMRs has not been evaluated, the proportions of patients with body mass index recorded in the EMRs has been found to be highly variable at different HMOs.[17]

Drug information typically pertains to dispensed prescriptions, as they come from prescription claims. Although records of prescriptions filled may provide more accurate measures of exposure over time than patient self-reports, they are not perfect measures of drug consumption. Nor do they provide full information on what was prescribed, since not all prescriptions are filled by patients. In addition, prescription medications filled out-of-plan, nonprescription medications, and those dispensed to hospital inpatients are not routinely captured in the health-plan dispensing files, and inexpensive prescriptions relative to the cost of deductibles may not be completely captured. Some health plans have restrictive formularies. However, different HMOs often have formularies that include different drugs; thus multisite projects often yield diverse drugs from a particular class. Newer agents may be somewhat slower to achieve widespread use than in the fee-for-service environment. Thus, evaluations of newer drugs for effectiveness and toxicity may be hampered.

Particular applications

Examples are drawn from the HMO Research Network (HMORN) and the Kaiser Permanente (KP) Comparative Effectiveness and Safety Research (CESR) initiative (Table 12.1), which are partially

Table 12.1 Affiliated managed care organizations and research departments

Managed care organization	Research department/ institute	Acronym	Location	HMORN	CESR
Geisinger Health System	Geisinger Center for Health Research	GHS	Pennsylvania	✓	
Group Health Cooperative	Group Health Research Institute	GHC	Washington State and Northern Idaho	✓	
Harvard Pilgrim Health Care	Harvard Pilgrim Health Care Institute and Harvard Medical School: Department of Population Medicine	HPHC	Massachusetts, New Hampshire and Maine	✓	
HealthPartners	HealthPartners Research Foundation	HPRF	Minnesota	✓	
Henry Ford Health System	Center for Health Services Research	HFHS	Michigan	✓	
Kaiser Permanente Colorado	Institute for Health Research	KPCO	Colorado	✓	✓
Kaiser Permanente Georgia	Center for Health Research—Southeast	KPG	Georgia	✓	✓
Kaiser Permanente Hawaii	Center for Health Research—Hawaii	KPH	Hawaii	✓	✓
Kaiser Permanente Mid-Atlantic	Department of Research	KPMAS	Maryland and Virginia		✓
Kaiser Permanente Northern California	Division of Research	KPNC	Northern California	✓	✓
Kaiser Permanente Northwest	Center for Health Research—Northwest	KPNW	Oregon and Washington	✓	✓
Kaiser Permanente Ohio	Division of Research	KPOH	Ohio		✓
Kaiser Permanente Southern California	Department of Research and Evaluation	KPSC	Southern California	✓	✓
Lovelace Health System	Lovelace Clinic Foundation	LCF	New Mexico	✓	
Maccabi Healthcare Services	Maccabi Institute for Health Services Research	MHS	Israel	✓	
Marshfield Clinic/ Security Health Plan of Wisconsin	Marshfield Clinic Research Foundation	MCRF	Wisconsin	✓	
Fallon Community Health Plan	Meyers Primary Care Institute	MPCI	Central Massachusetts	✓	
Scott and White Health System	Scott and White Division of Research and Education	S&W	Texas	✓	

overlapping consortia of health plans which exemplify many characteristics of HMOs described above. The health plans that comprise these consortia serve large, geographically and ethnically diverse defined populations, have varied health-care delivery models and practice patterns, automated claims data, access to full medical records, access to providers, and the ability to work with the health-plans' members. The 16 member health plans of the HMORN, 15 in the US and one in Israel (Tables 12.1, 12.2; Figure 12.1) include nearly 13 million current enrollees, enough to address many questions that could not be addressed with any individual plan's population alone. A wide array of medical-care delivery models is represented in the HMORN, including staff, group and network, as well as independent physicians associations (IPAs, organizations that contract with HMOs on behalf of physicians who are not employed by the HMO). The KP CESR initiative, established in 2009, is a culturally and geographically diverse, distributed research network spanning nine US states and the District of Columbia. It is comprised of all eight KP regions, six of which are also HMORN members.

Historical context

Early pharmacoepidemiologic studies were initiated at these managed care organizations over 40 years ago when Kaiser Permanente Northern California received a contract from the US Food and Drug Administration (FDA) to develop one of the first computerized systems for monitoring adverse drug reactions in both inpatients and outpatients.[18,19] The project was successful in compiling databases containing outpatient diagnoses and prescriptions dispensed to over 217 000 individuals between 1969 and 1973.[19] Subsequently, a two-phase surveillance program was developed and funded by the US National Cancer Institute (NCI) to monitor possible carcinogenic effects of drugs. In the hypothesis-seeking phase of that study, the 143 574-person cohort with pharmacy exposure data has been followed for more than 25 years.[20–23] In exploratory analyses, incidence of 56 types of cancer has been assessed in age- and sex-adjusted comparisons of users with non-users of each of 215 drugs or drug groups. In the hypothesis-testing phase of that study, selected associations were re-examined in case–control studies that were better able to adjust for potential confounding factors. One positive finding was an association of barbiturate use with lung cancer.[24,25] This data source has also been an important source of evidence for the lack of association for drugs suspected of causing cancer.[26–28]

Standard data layouts

As noted above, the adoption of standard data layouts that use a common data model has facilitated multi-HMO research. The existence of data files that adhere to a common data model allows creation of libraries of reusable data management and analysis programs. These standard layout files also greatly facilitate multicenter research, since it is possible for a single program to be used in multiple sites with negligible modification. This practice improves the efficiency and consistency of multicenter research.

The core data resource for research in the HMO Research Network is its Virtual Data Warehouse (VDW).[29,30] The VDW is "virtual" in the sense that it is a distributed research dataset, with the data transformed according to a common data model remaining at each site (Figure 12.2). Within each research center, data are extracted from the extensive, local health-plan data systems into the site-specific VDWs and configured into 14 tables using standard variable names and values (Figure 12.3). The primary content areas of the health-plan data systems include enrollment, demographics, outpatient prescription drug dispensings, outpatient diagnoses and procedures, utilization of clinical services, hospitalizations, geocoding of member addresses, and tumor characteristics (for cancer patients). Other content areas include vital signs, laboratory results, and death.

For each content area, a data dictionary specifies the format for each element including variable name, variable label, extended definition, code values, and value labels. Local site programmers have mapped and transformed data elements from their local data systems into the standardized set of variable definitions, names, and codes. The extract, transform, and load (ETL) procedure produces SAS

Table 12.2 Demographic characteristics of HMO Research Network member health plans (primary model / % EMR data)

Health plan	GHS	GHC	HPHC	HPRF	HFHS	KPCO	KPG	KPH	KPNC	KPNW	KPSC	LCF	MHS	MCRF	MPCI	S&W
Year established	1915	1947	1969	1957	1948	1969	1985	1958	1945	1942	1947	1973	1941	1916	1977	1982
Primary model	Mixed	HMO	Mixed	Mixed	Mixed	HMO	Mixed	HMO	HMO	HMO	HMO	HMO	HMO	Mixed	Mixed	HMO
Total enrolled, x1000	229	617	762	687	208	451	271	216	3,130	471	3,324	194	1,800	160	220	203
% with EMR data	36	69	35	83	TBD	83	100	98.5	TBD	98	TBD	100	100	100	60	75
Age																
% ≤17 years	19	20	24	35	18	22	24	22	22	23	25	39	41	24	19	30
% 18–44 years	29	33	39	28	29	34	39	35	35	34	36	25	32	33	33	24
% 45–64 years	28	33	33	30	35	30	31	30	29	31	28	20	19	26	29	28
% 65 +	24	13	4	4	18	14	7	13	13	13	11	15	5	17	19	9
Gender																
% Female	52	53	52	52	55	53	53	50	52	52	52	55	52	52	52	53
Race*																
% White	96	82	75	81	54	74	63	25	51	84	38	55	95	97	87	84
% African American	<1	3	16	9	33	5	33	<1	8	3	8	1	0	<1	2	8
% Asian American	0	6	5	5	3	3	<1	63	17	5	10	1	0	<1	3	2
% American Indian	<1	1	<1	1	<1	1	<1	<1	<1	1	<1	2	0	<1	<1	1
% Hispanic	1.4	4	4	2	1	15	<1	3	19	6	41	38	0	<1	8	7
% Other	<1	3	0	2	8	2	4	17	5	1	<1	0	5	<1	0	7
Member retention (% still enrolled at 1, 3, 5 years; 2003–2008)																
% enrolled at 1 year	82	84	78	73	99.5	83	87	85	87	82	87	78	99	92	95	81
% enrolled at 3 years	54	66	47	46	86.4	66	67	72	75	66	70	66	98	78	92	81
% enrolled at 5 years	41	55	35	38	62.6	56	54	63	66	57	59	57	98	68	92	62

* May be >100% if multiple responses allowed at collection, "other" may include persons reporting multiple races.
See Table 12.1 for definitions of abbreviations.

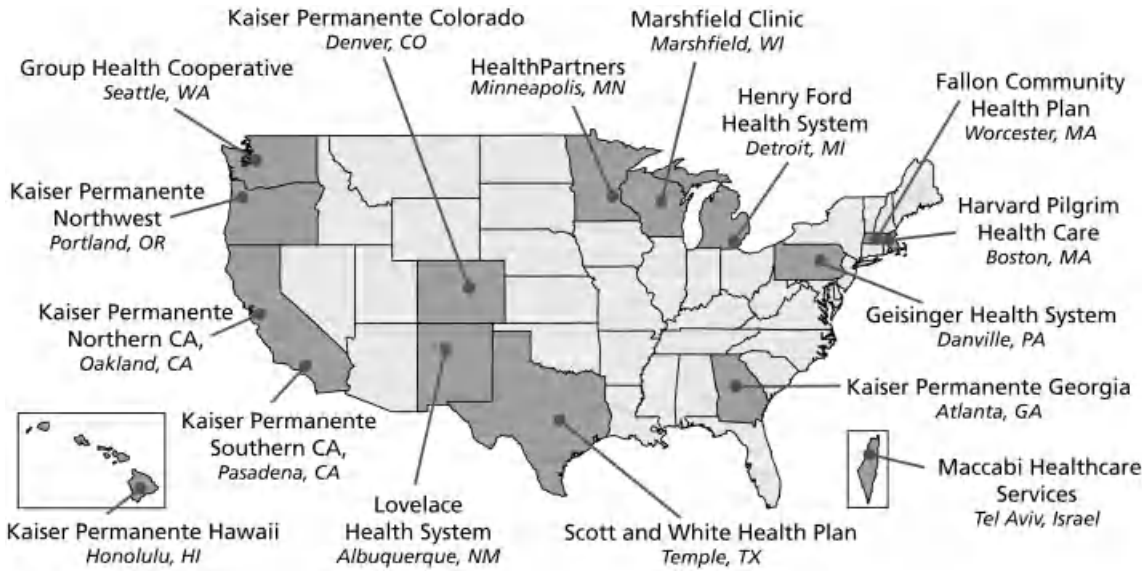


Figure 12.1 The HMO research network.

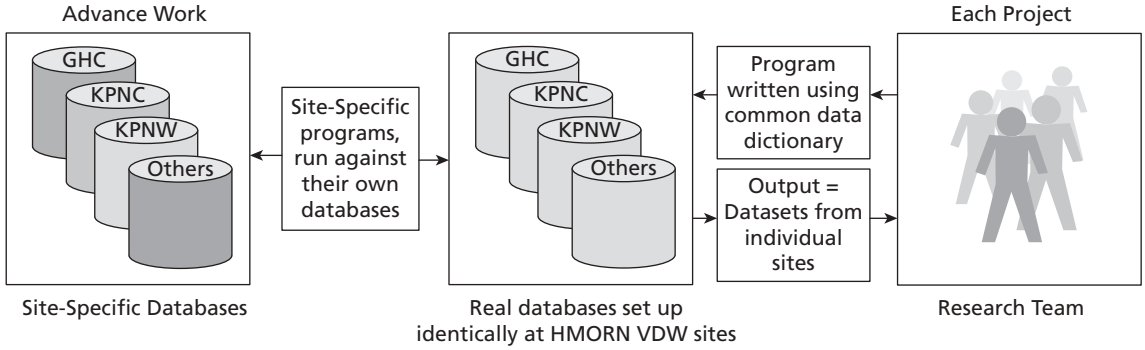


Figure 12.2 The HMO Research Network virtual data warehouse. GHC, Group Health Cooperative; KPNW, Kaiser Permanente Northwest; KPNC, Kaiser Permanente Northern California.

datasets with a common structure at each HMO, thereby allowing an analyst at one site to write one SAS program—based on the VDW's data dictionary—which is then distributed to extract and/or analyze data at participating sites with minimal edits by the local analyst (Figure 12.2).[31] Output (usually de-identified individual-level or aggregate-level data files) is then securely transferred to the research center leading the specific project where analytic datasets are assembled from these files. The VDW greatly reduces and standardizes the preparatory work needed to assemble cohorts, count events, and capture exposure and co-morbid conditions across systems.

HMORN sites have developed a general purpose "Collaboration Toolkit" to assist in planning and carrying out multisite research. Resources pertinent to data development include multicenter IRB

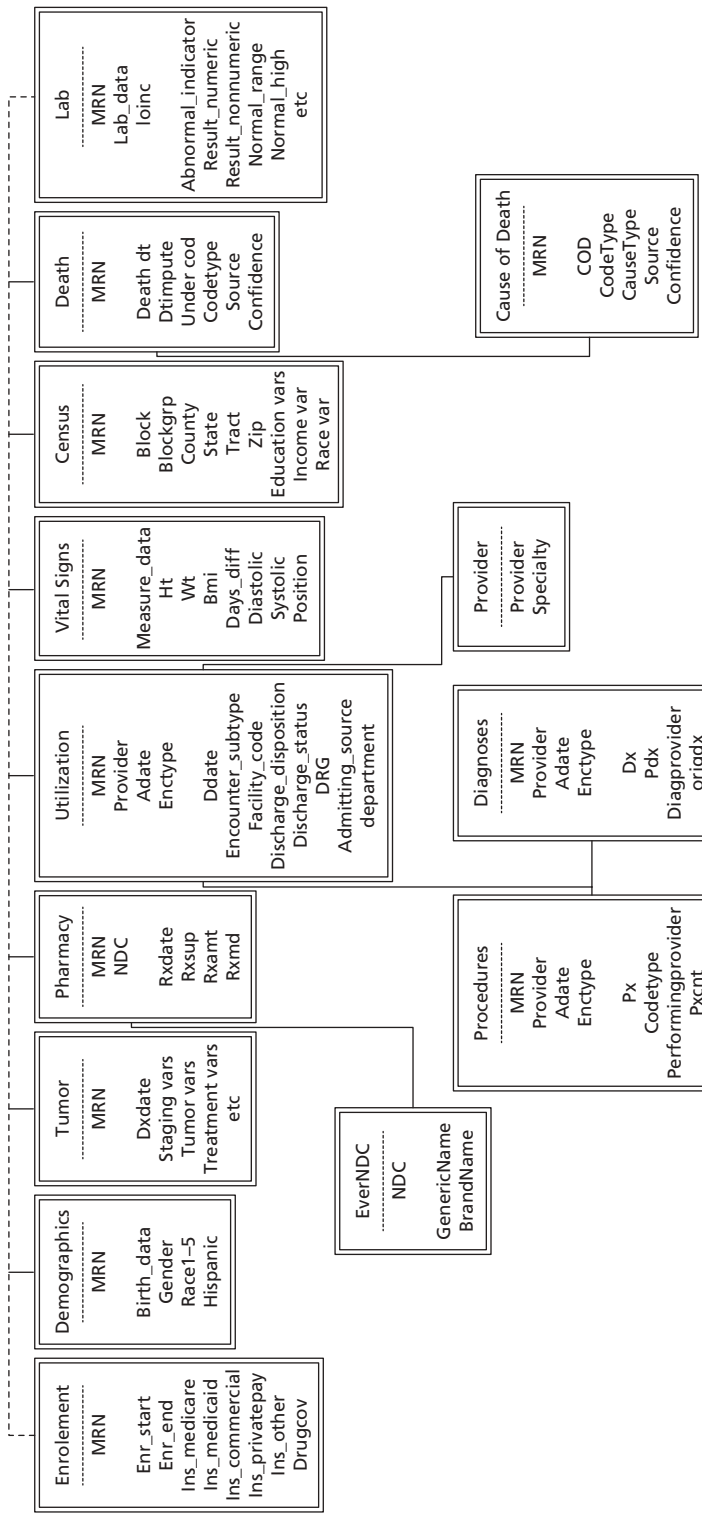


Figure 12.3 Virtual data warehouse structure. MRN, medical record number; NDC, national drug code.

review, multisite subcontract, and data use agreement (DUAs) descriptions and instructions (http://www.hmoresearchnetwork.org).

Examples of multisite, multiproject research programs using HMO data within HMORN and KP CESR

The formal collaborations and research programs across the HMORN and KP CESR span multiple therapeutics topics that are funded by the FDA, the US Centers for Disease Control and Prevention (CDC), the US Agency for Healthcare Research and Quality (AHRQ), the US National Institutes of Health (NIH), non-profit foundations, and private organizations. These programs include:

- HMORN Center for Education and Research on Therapeutics (HMORN CERT; supported by AHRQ);
- Developing Evidence to Inform Decisions about Effectiveness (DEcIDE; AHRQ) center;
- Cancer Research Network (CRN; NCI);
- Cardiovascular Research Network (CVRN; US National Heart, Lung, and Blood Institute, [NHLBI]);
- Mental Health Research Network (MHRN; US National Institute of Mental Health, [NIMH];)
- Scalable PArtnering Network for Comparative Effectiveness Research (SPAN; AHRQ);
- Distributed Research Network (DRN; AHRQ HMORN DEcIDE center);
- contracts from the FDA's Center for Drug Evaluation and Research (CDER) and its Center for Biologics Evaluation and Research (CBER), and several other newly emerging collaboratives;
- the Vaccine Safety Datalink (VSD; CDC) is comprised entirely of HMORN member health plans with defined populations (also see Chapter 26); and
- most of these health plans are active participants in the FDA's Mini-Sentinel Program under the leadership of the Harvard Pilgrim Health Care (HPHC).

HMORN Center for Education and Research in Therapeutics (CERT)

The HMORN CERT (Table 12.2) is one of 14 CERTs created in response to the FDA Modernization Act of 1997. The mission of the CERTs includes research and education to advance the optimal use of drugs, medical devices, and biological products (http://certs.hhs.gov/). The CERTs program is funded and run as a cooperative agreement by AHRQ in consultation with the FDA.

The HMORN CERT's theme is to assess the use, effectiveness, and safety of therapeutics in the population laboratories defined by the members, clinicians, and data systems of its health plans. The investigators accomplish this by taking advantage of the research and dissemination opportunities afforded by health plans' defined populations, their large provider groups, and their unique data sources.[32] These studies use the health plans' defined populations, delivery systems and patient populations to enhance both the study design and the generalizability of study findings.

HMORN CERT investigators have conducted a wide variety of drug safety studies that range from classic drug safety studies such as the association between alendronate and risk of gastrointestinal perforation or bleeding[33] to developing and testing methods for early detection of adverse drug events via active surveillance.[34,35] These latter studies, involving nine HMOs, concluded that sequential analysis methods originally developed for surveillance of vaccines were also useful for drugs.

Clinical trials conducted by the CERT have employed innovative designs, such as randomization of entire medical practices to different management strategies. Other CERT studies use direct outreach to health-plan members. A cluster randomized study of direct-to-patient outreach evaluated the ability of mailed intervention to increase persistence of use of beta-blocker medication after acute myocardial infarction (AMI). Patients identified from hospital discharge diagnoses as having had a recent AMI were randomized at the clinic level to intervention (mailed information) or usual care. Outcomes of persistency of use were ascertained directly from pharmacy data. For every 16 patients who received the intervention, one additional patient became adherent, defined as having at least an 80% proportion of days covered (PDC) during the year following the AMI.[36]

Table 12.3 Representative pharmacoepidemiologic studies from HMO-based research

Title	First author/ citation
Arthritis and non-traumatic joint disorders	
Use of anti-inflammatory drugs and lower esophageal sphincter-relaxing drugs and risk of esophageal and gastric cancers	Fortuny J[113]
The role of cigarette smoking and statins in the development of postmenopausal osteoporosis: a pilot study utilizing the Marshfield Clinic Personalized Medicine Cohort	Giampietro PF[114]
Medication adherence of patients with selected rheumatic conditions: a systematic review of the literature	Harrold LR[115]
Cancer	
Diffusion of aromatase inhibitors for breast cancer therapy between 1996 and 2003 in the Cancer Research Network	Aiello EJ[60]
Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research	Aiello Bowles EJ[116]
Cardiovascular medication use and risk for colorectal cancer	Boudreau DM[64]
Impact of hormone therapy on false-positive recall and costs among women undergoing screening mammography	Boudreau DM[117]
Estrogen receptor genotype is associated with risk of venous thromboembolism during tamoxifen therapy	Onitilo AA[110]
Cardiovascular disease	
Current use of opposed estrogen and unopposed estrogen and the risk of acute myocardial infarction among women with diabetes: the Northern California Kaiser Permanente Diabetes Registry 1995–1998	Ferrara A[118]
Comparative effectiveness of β-adrenergic antagonists (atenolol, metoprolol tartrate, carvedilol) on the risk of rehospitalization in adults with heart failure	Go AS[8]
Comparative effectiveness of different β-adrenergic antagonists on mortality among adults with heart failure in clinical practice	Go AS[9]
Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs	Graham DJ[88]
Antihypertensive treatment with ACE inhibitors or beta-blockers and risk of incident atrial fibrillation in a general hypertensive population	Heckbert SR[10]
Depression and other mental health disorders	
Does antidepressant adherence have an effect on glycemic control among diabetic antidepressant users?	Bambauer KZ[119]
Suicide risk in bipolar disorder during treatment with lithium and divalproex	Goodwin FK[107]
Modeling the impact of enhanced depression treatment on workplace functioning and costs: a cost–benefit approach	Lo Sasso AT[120]
Cost-effectiveness of systematic depression treatment among people with diabetes mellitus	Simon GE[121]
Suicide attempts among patients starting depression treatment with medications or psychotherapy	Simon GE[122]
Effects of a limit on Medicaid drug-reimbursement benefits on the use of psychotropic agents and acute mental health services by patients with schizophrenia	Soumerai SB[123]
Diabetes mellitus	
A cohort study of the incidence of serious acute liver injury in diabetic patients treated with hypoglycemic agents	Chan KA[124]
Initial non-adherence, primary failure and therapeutic success of metformin monotherapy in clinical practice	Nichols GA[125]
Improving medication adherence: challenges for physicians, payers, and policy makers	O'Connor PJ[126]
Treatment intensification and risk factor control: toward more clinically relevant quality measures	Selby JV[127]

Table 12.3 (*Continued*)

Title	First author/ citation
Infectious disease including HIV/AIDS	
Statin use and risk of community acquired pneumonia in older people: population based case-control study	Dublin S[128]
Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project	Greene SK[129]
Risk of medically attended local reactions following diphtheria toxoid containing vaccines in adolescents and young adults: a Vaccine Safety Datalink study	Jackson LA[130]
Compliance with multiple-dose vaccine schedules among older children, adolescents, and adults: results from a vaccine safety datalink study	Nelson JC[131]
Increased use of second-generation macrolide antibiotics for children in nine health plans in the United States	Stille CJ[132]
Pulmonary disease/ asthma	
Intranasal steroids and the risk of emergency department visits for asthma	Adams RJ[133]
Clinical effectiveness research in managed-care systems: lessons from the Pediatric Asthma Care PORT	Finkelstein JA[134]
Asthma drug use and the development of Churg–Strauss syndrome (CSS)	Harrold LR[135]
Pain	
Trends in long-term opioid therapy for chronic non-cancer pain	Boudreau D[136]
Opioid prescriptions for chronic pain and overdose: a cohort study	Dunn KM[137]
De facto long-term opioid therapy for non-cancer pain	Korff MV[138]
Pain management in the last six months of life among women who died of ovarian cancer	Rolnick SJ[139]
Trends in prescribed opioid therapy for non-cancer pain for individuals with prior substance use disorders	Weisner CM[140]
Pregnancy including preterm birth	
Antidepressant medication use and risk of persistent pulmonary hypertension of the newborn	Andrade SE[82]
Use of antidepressant medications during pregnancy: a multisite study	Andrade SE[84]
Risks of congenital malformations and perinatal events among infants exposed to antidepressant medications during pregnancy	Davis RL[86]
Evaluation of gestational age and admission date assumptions used to determine prenatal drug exposure from administrative data	Raebel MA[141]
Methods	
Active influenza vaccine safety surveillance: potential within a healthcare claims environment	Brown JS[142]
Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations	Brown JS[35]
Early detection of adverse drug events within population-based health networks: application of sequential testing methods	Brown JS[34]
Active surveillance of vaccine safety: a system to detect early signs of adverse events	Davis RL[70]
Practical clinical trials for translating research to practice: design and measurement recommendations	Glasgow RE[143]
Modeling the cumulative risk of a false-positive screening test	Hubbard RA[144]
Real-time vaccine safety surveillance for the early detection of adverse events	Lieu TA[76]
Potential population-based electronic data sources for rapid pandemic influenza vaccine adverse event detection: a survey of health plans	Moore KM[145]
Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals	Xu S[50]

(*Continued*)

Table 12.3 (*Continued*)

Title	First author/ citation
Monitoring high-risk medication therapy	
Improved therapeutic monitoring with several interventions: a randomized trial	Feldstein AC[46]
Academic detailing to improve laboratory testing among outpatient medication users	Lafata JE[146]
Evaluation of laboratory monitoring alerts within a computerized physician order entry system for medications orders	Palen TE[47]
Diabetes and drug-associated hyperkalemia: effect of serum potassium monitoring	Raebel MA[49]
Improving laboratory monitoring at initiation of drug therapy in ambulatory care: a randomized trial	Raebel MA[44]
Monitoring of drugs with a narrow therapeutic range in ambulatory care	Raebel MA[37]
Prescribing safety	
Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care	Brown JS[31]
Potential drug-drug interactions in the outpatient setting	Lafata JE[147]
Randomized trial to improve prescribing safety during pregnancy	Raebel MA[148]
Randomized trial to improve prescribing safety in ambulatory elderly patients	Raebel MA[149]
The impact of prescribing safety alerts for elderly persons in an electronic medical record: an interrupted time series evaluation	Smith DH[55]

The CERT has also evaluated clinicians' practices in monitoring drug therapy. Periodic laboratory monitoring is recommended for drugs that carry a risk of organ system toxicity or electrolyte imbalance, but adherence to these recommendations was incompletely characterized. Raebel and colleagues conducted a series of studies involving ten HMOs sites to evaluate laboratory monitoring among patients dispensed these high-risk medications and to assess the patient correlates of laboratory monitoring.[37–42] Across all high-risk drugs evaluated, important proportions of patients had not received recommended monitoring at initiation of, or during ongoing, therapy. For example, 39% of individuals beginning therapy with one of these medications in the timeframe of the study did not have indicated baseline laboratory monitoring.[42] Medical record reviews documented that the administrative records were accurate in the majority of situations (i.e. 72–89%). This series of studies demonstrates the utility of linking pharmacy and laboratory administrative databases to evaluate quality of care questions. These results informed development of the National Committee for Quality Assurance (NCQA) parameters for annual monitoring for patients on persistent medications. These studies also drove targeted intervention studies that improved laboratory monitoring rates as well as observational studies that confirmed patients with monitoring had a lower risk of adverse events than those without monitoring.[43–50]

The CERT evaluated real-time decision support via computerized prescriber order entry systems, evaluating medication safety alerts delivered at the time of prescribing via the EMR.[51–55] The alerts targeted prescribing for the elderly, renal dosing, and drug interactions and allowed prescribing clinicians to change medication orders to a preferred agent. The study included a randomized intervention measuring the incremental effect of a group academic detailing effort wherein clinicians were randomized to receive an educational session on medication safety. The alerts and detailing efforts were informed by preliminary qualitative work assessing clinician barriers to using EMR-based alerts, and preferred modes of education.[56]

Cancer Research Network (CRN)

The CRN consists of the research programs, enrolled populations, and data systems of 14 HMORN sites nationwide (http://crn.cancer.gov/), originally funded by the NCI in 1998. Its goal is to conduct

research on cancer prevention, early detection, treatment, long-term care, surveillance, and cancer communication and dissemination and implementation research (http://crn.cancer.gov/projects/projects.php). The CRN is equipped to study cancer control at the patient, provider, and system levels, and has experience with a range of data collection strategies. CRN activities have generated more than 140 journal publications in a range of disciplines. Studies include evaluations of the patterns and trends in use of hormone therapy,[57,58] tamoxifen,[59] and other chemotherapy;[59–62] assessments of the risk of cancer associated with exposure to statins and other cardiovascular medications;[63–66] and an evaluation of the impact of a decision aid designed to educate women about their risk of breast cancer and the risks and benefits of tamoxifen.[67]

Cancer Care Outcomes Research and Surveillance Consortium (CanCORS)

The CRN Cancer Care Outcomes Research and Surveillance Consortium (CanCORS), involving five HMORN sites, is a part of the larger CanCORS national collaboration of eight grantees funded by the NCI in collaboration with the Department of Veteran Affairs. The goal of CanCORS is to measure the quality of cancer care and associated health outcomes by evaluating patterns of treatment, decision-making, and outcomes for lung and colorectal cancers. Several structural aspects of the participating HMORN sites enhance their contributions to CanCORS, including natural ties to patients; access to accurate demographic, contact and medical information before contacting participants; and the cooperation and strong support of health plans' medical groups in encouraging patients to join the study. Data from patient surveys will be supplemented with medical record data to evaluate cancer treatments recommended and received, adverse events associated with treatment, and clinical outcomes, as well as patients' concomitant diagnoses, diagnostic tests, and staging evaluation.[68] Among the aims, CanCORS is investigating racial, ethnic, and socioeconomic differences in cancer care, along with differences by age group.

Cancer Communications Research Center (CCRC)

The Cancer Communications Research Center (CCRC) is funded by the NCI and is considered a CRN-Affiliated Center. CCRC brings together six of the HMORN sites to identify and describe optimal communication structures and processes in organizations that facilitate patient-centered communication in cancer care. The Center's projects include: Testing an Optimal Model of Patient-Centered Cancer Care, Improving Physician-Parent Communication to Reduce Home Medication Errors and Improve Adherence in Children, and Effective Communication for Preventing and Responding to Oncology Adverse Events.

Developing Evidence To Inform Decisions About Effectiveness (DEcIDE) Network, the Diabetes Multi Center Research Consortium (DMCRC), and the Distributed Research Network (DRN)

The DEcIDE network is part of AHRQ's Effective Health Care program (http://effectivehealthcare.ahrq.gov), created in response to the Medicare Modernization Act of 2003. DEcIDE focuses primarily on comparative effectiveness of therapies, as assessed from observational studies and practical clinical trials in "real-world" populations. Examples of DEcIDE task orders awarded to the HMORN include: (i) observational studies of comparative effectiveness of beta-blockers for heart failure on hospital readmission and mortality;[8,9] (ii) development of evidence and educational/system approaches to reduce prenatal exposure to medications with a potential for fetal harm; and (iii) a cluster randomized trial in 45 hospitals of three different methods to reduce acquisition of methicillin resistant *S. aureus* in intensive care units.[69] While this latter study is an intervention, it is grounded in the use of observational data to identify and characterize the cohorts, and to assess the outcomes of their treatment.

The HMORN DEcIDE Center Diabetes Multi Center Research Consortium (DMCRC) includes 12 HMORN health plans plus external partners. The DMCRC is developing a comprehensive comparative

effectiveness research agenda, and a distributed research database of patients with diabetes. Examples of DMCRC studies include: (i) cohort studies comparing diabetes patients undergoing bariatric surgery with similar patients receiving usual clinical care, and (ii) outcomes of further intensifying diabetes therapy in patients already on at least two oral medications or basal insulin to maintain "tight" glycemic control (A1c < 7%). This study attempts to address the apparent lack of benefit on cardiovascular disease endpoints of tight glycemic control observed in recent trials.

AHRQ has also funded the HMORN DEcIDE center to create new capabilities for distributed studies of the safety and effectiveness of treatments, including drugs, vaccines, and medical devices. The program, PopMedNet (www.popmednet.org), allows data partners to create both large and small data networks, while retaining control over the uses and users of their data resources.[31] PopMedNet is an efficient, reusable infrastructure through which routinely collected health-care data and related information can be assembled and analyzed to support decision making by patients, providers, and policy makers. Participating investigators can distribute queries through network software, execute queries against local data, and return aggregated results to the end user. The software is capable of supporting a variety of study types, including observational studies, quasiexperimental studies, clinical trials, and registries.

Vaccine Safety Datalink (VSD)

The CDC-funded VSD (also see Chapter 26) is comprised of eight HMORN sites. The VSD was established in 1991 to monitor immunization safety and address the gaps in scientific knowledge about rare and serious events following immunization (http://cdc.gov/vaccinesafety/Activities/VSD.html). Currently, VSD collects vaccination and comprehensive medical information on nearly 9 million members from these HMOs annually (3% of the US population). While most of the VSD data sources are identical to the HMORN's Virtual Data Warehouse, the vaccine exposure data is obtained from the HMOs' electronic medical records. The VSD uses the distributed data approach described above. The VSD has developed a standardized approach for performing near-real-time safety surveillance of new vaccines as these vaccines enter the US market.[70] VSD Rapid Cycle Analysis (RCA) studies rely on data automatically updated weekly and sequential statistical analyses to compare the rate of occurrence of prespecified adverse events following receipt of a vaccine with expected rates among persons not exposed to the vaccine. The success of the VSD RCA studies has led to the use and extension of this approach to the surveillance of drug safety in the CERT, described above, and the FDA mini-Sentinel program, described below. Examples of VSD studies include the following.

Ritzwoller *et al.*[71] used survival analysis methods to show that a vaccine that was not well matched to the circulating influenza A strain retained some effectiveness in preventing pneumonia and influenza-like illness among children. VSD researchers also published three landmark papers evaluating the effects of unrecognized confounding on the association between influenza vaccination and mortality.[72–74] The results of this research agenda provided support for the increasing recognition of an alternate explanation—that the apparent vaccine effect is due, at least in part, to preferential selection of vaccination by healthier persons who are at relatively low risk of death during the winter months. Their work also emphasizes the need for rich sources of clinical data such as EMR to reach beyond standard adjustment of confounding using encounter diagnosis codes to specific measures of disease severity such as functional status.[73]

In a study of a well-defined cohort with verified individual-level vaccination data, Glanz and colleagues examined the relationship between parental refusal of pertussis vaccination and risk of pertussis infection in children.[75] They found that vaccine refusers had a 23-fold increased risk (OR 22.8, 95% CI = 6.7 – 77.5) for pertussis when compared with vaccine acceptors. A full 11% of pertussis cases in the entire study population were attributed to vaccine refusal.

Finally, the VSD has established a program of active surveillance of new vaccines, to identify preselected adverse reactions at the earliest

possible time. Davis *et al*. led early work demonstrating the concept of using historical data from four health plans participating in the VSD to detect an increase in intussusception after Rotashield (rotavirus vaccine).[70] Subsequent work led to the development of the maximized sequential probability ratio test (maxSPRT) and evaluation of its application to assessment of vaccines and, later, to drugs.[34,76] The VSD now routinely, prospectively evaluates new vaccines. This program detected an increased risk of febrile seizures following measles–mumps–rubella–varicella vaccine, leading the Advisory Committee on Immunization Practices to change its recommendation regarding the use of this vaccine.[77]

Cardiovascular Research Network (CVRN)

The CVRN, funded by NHLBI since 2007 (http://www.cvrn.org), includes 15 HMORN sites. It has created a framework to address questions about contemporary cardiovascular epidemiology, optimal management, and associated clinical outcomes within large community-based populations where most clinical care is delivered. Examples of CVRN research projects include studies of: (i) hypertension diagnosis, management, and control; (ii) quality of care and outcomes of warfarin therapy for atrial fibrillation and venous thromboembolism; and (iii) the use and outcomes of implantable cardioverter defibrillators for prevention of sudden death. The network also established capacity to rapidly respond to emerging cardiovascular disease research questions and to facilitate external collaborations.

CVRN has reported trends in therapy[78] and diagnosis[79] of hypertension. The CVRN has also evaluated the comparative effectiveness of second-line therapies in patients whose blood pressure was not controlled with first-line therapy.[80] The study was conducted using patients enrolled in the CVRN Hypertension Registry at three HMORN sites. Patients whose blood pressure was not controlled with a thiazide diuretic were followed after addition of either an angiotensin-converting enzyme (ACE) inhibitor or beta-blocker. Blood pressure control was comparable for the two agents. In addition, the rates of incident myocardial infarction and stroke were also similar during an average of 2.3 years of follow-up.

Medication Exposure In Pregnancy Risk Evaluation Program (MEPREP)

The FDA-funded Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP), includes ten HMORN sites, plus Tennessee Medicaid. Its purpose is to study the effects of prescription medications used during pregnancy. To overcome the challenges presented by the lack of clinical trial data about the use of medications during pregnancy, MEPREP links health care information for mothers and their babies. Collectively, the participating sites have health care information for about one million births from 2001–2007. Examples of studies conducted in these environments include: (i) assessment of medication use during pregnancy and birth outcomes, and (ii) the effects of antidepressant medications and cardiovascular medications on birth defects and perinatal outcomes.[81–87]

FDA Mini-Sentinel Program

The FDA's Mini-Sentinel program (http://www.minisentinel.org) is a major component of the FDA's Sentinel Initiative (see also Chapter 30). The Mini-Sentinel program will create a "laboratory" for developing and evaluating safety surveillance scientific methods and offers FDA the opportunity to evaluate medical product safety in existing automated health-care data systems while learning about the barriers and challenges inherent in these activities. This consortium, led by the Harvard Pilgrim Health Care Institute, involves most HMORN sites, plus large national health plans, free-standing registries, and hospitals.

FDA contracts: epidemiologic studies of adverse effects of marketed drugs

Nearly all HMORN sites participate in the FDA's Pharmacoepidemiologic Research Program, which assists the FDA's Center for Drug Evaluation and Research (CDER) Office of Surveillance and Epidemiology. This program and its predecessors have evaluated both direct adverse effects of drugs, such as rhabdomyolysis associated with lipid lowering agents,[88] and appropriateness of medication

use.[89] The latter study contributed to the withdrawal of cisapride.

Mental Health Research Network (MHRN)

The MHRN, funded by NIMH, includes nine HMORN health plans. Its portfolio includes studies of: (i) practice variation in high- and low-value care for mood disorders, (ii) a geographically and ethnically diverse autism registry for effectiveness studies, and (iii) a longitudinal analysis of selective serotonin reuptake inhibitor warnings and suicide in youth.

Scalable PArtnering Network (SPAN) for Comparative Effectiveness Research

The SPAN for Comparative Effectiveness Research (CER): Across Lifespan, Conditions, and Settings (AHRQ) is a distributed research network intended to conduct large CER studies using data collected on patient-reported outcomes. SPAN's cohorts include: (i) a cohort of children with attention deficit hyperactivity disorder (ADHD), with and without identified learning and developmental disorder who have received differing treatments, and (ii) a cohort of adults with obesity, with and without diagnosed depression, who were and were not treated with weight-loss surgery.

Additional pharmacoepidemiologic research studies conducted using HMO data

Databases from these HMOs are frequently used to examine patterns of drug prescribing and adherence,[90–95] utilization,[93,96] costs, and cost-effectiveness associated with use of specific therapies.[97,98] For example, data have been used to describe utilization of antipsychotics,[99] antidepressants,[100] mood stabilizers,[101] and anxiolytic medications.[102] Further, these data have had an impact on regulatory decisions such as labeling changes for triazolam[103] and generated important methods for use in pharmacoepidemiology.[102,104]

Keith, Smith, *et al.*[105,106] used computerized laboratory results to identify a patient population with chronic kidney disease (CKD) based on the US National Kidney Foundation's staging guidelines. Interestingly, most of the patients identified did not carry a diagnosis of CKD in their medical record. Patients with CKD were found to have health-care costs twice that of age- and sex-matched patients without CKD. Potential under-treatment was identified in this population in that, even in the most severe stage of disease, only about half of those with anemia (defined as a hemoglobin level less than 12 g/dL with normal mean corpuscular volume) were treated.

Suicide is a frequent consequence of bipolar disorder. Concern has been expressed regarding the relative effectiveness of current therapeutic options for reducing suicide risk in this condition. In a cohort study of 20 638 patients with diagnosed bipolar disorder conducted at KPNC and Group Health Cooperative (GHC),[107] risk in lithium users was found to be substantially and significantly lower than that for users of divalproex. Findings persisted with adjustment for co-morbidities and other current treatments. Interestingly, the authors also identified the increased risk for suicide in those persons initiating a new therapy, whether switching from lithium to divalproex or vice versa.

Lastly, HMORN and CESR sites are involved in innovative personalized medicine, biobanking, and population-based pharmacogenomics research. At Kaiser Permanente Northern California, the Research Program on Genes, Environment, and Health is one of the largest research projects in the USA to examine the genetic and environmental factors that influence many common diseases. The completed biobank resource will link together electronic medical records, data on behavioral and environmental factors, and biobank data from 500 000 consenting health-plan members.

Marshfield Clinic Research Foundation's Personalized Medicine Research Project has undertaken numerous pharmacogenomics research projects.[108–112] Some of these include: (i) polymorphisms in cytochrome b5 and its reductase and the risk of sulfonamide hypersensitivity; (ii) how genetic differences can influence an individual's response to statins;[111,112] (iii) genetic variation and response to metformin therapy; (iv) genetic predictors of angiotensin-converting enzyme inhibitor (ACEi)-associated angioedema; and (v) the association between polymorphisms in estrogen receptor

genes and clinical outcomes in breast cancer patients receiving tamoxifen treatment.[110]

Another example of pharmacogenomics work is a project with the International Serious Adverse Event Consortium (iSAEC), a non-profit organization comprised of pharmaceutical companies, the Wellcome Trust, and academic institutions, with input from the FDA. The mission of the iSAEC is to identify DNA-variants useful in predicting the risk of drug-related serious adverse events (SAEs). The current collaboration is addressing genetic variants and drug-induced liver injury, Stevens–Johnson syndrome, and extreme weight gain among users of atypical antipsychotic medications.

The future

HMO-based research is likely to evolve in several ways. These organizations are well suited to use cluster randomization, at the level of practices or health plans, to evaluate treatment strategies that are in equipoise and which cannot be evaluated through purely observational methods. As described above, the health plans' observational data makes such studies especially efficient and economical, while providing highly relevant real-world data on comparative clinical effectiveness. HMOs can also likely improve the efficiency of more conventional clinical trials, by using health-plan data to obtain preliminary estimates of the number of potential eligible subjects and to support identification, recruitment, and follow-up of study participants.

HMOs are also relatively early adopters of electronic health records, and thus likely to be able to incorporate these data more quickly into pharmacoepidemiologic studies. The use of natural language processing methods to allow more effective use of electronic health records' free text will also improve the depth and quality of data available. Additional initiatives that are likely to grow in importance include more linkage to external data sources, such as state immunization registries. HMOs are also well positioned to collect biological specimens, either for immediate use or to store in a specimen bank to support later research. HMOs are also reasonably well positioned to engage their members to obtain historical and behavioral information that is not routinely collected in health records.

References

1. Hazlehurst B, Frost HR, Sittig DF, Stevens VJ. MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. *J Am Med Inform Assoc* 2005; **12**: 517–29.
2. Hazlehurst B, Mullooly J, Naleway A, Crane B. Detecting possible vaccination reactions in clinical notes. *AMIA Annu Symp Proc*, 2005, p. 306–10.
3. Hazlehurst B, Naleway A, Mullooly J. Detecting possible vaccine adverse events in clinical notes of the electronic medical record. *Vaccine* 2009; **27**: 2077–83.
4. Hazlehurst B, Sittig DF, Stevens VJ, Smith KS, Hollis JF, Vogt TM, *et al*. Natural language processing in the electronic medical record: assessing clinician adherence to tobacco treatment guidelines. *Am J Prev Med* 2005; **29**: 434–9.
5. Pardee R. PS2-01: a system for computer-assisted rapid case ascertainment of breast, lung and colorectal cancers. *Clin Med Res* 2010; **8**: 186.
6. Smith DH, Johnson ES, Russell A, Hazlehurst B, Muraki C, Nichols GA, *et al*. Lower visual acuity predicts worse utility values among patients with type 2 diabetes. *Qual Life Res* 2008; **17**: 1277–84.
7. Go AS, Lee WY, Yang J, Lo JC, Gurwitz JH. Statin therapy and risks for death and hospitalization in chronic heart failure. *JAMA* 2006; **296**: 2105–11.
8. Go AS, Yang J, Gurwitz JH, Hsu J, Lane K, Platt R. Comparative effectiveness of beta-adrenergic antagonists (atenolol, metoprolol tartrate, carvedilol) on the risk of rehospitalization in adults with heart failure. *Am J Cardiol* 2007; **100**: 690–6.
9. Go AS, Yang J, Gurwitz JH, Hsu J, Lane K, Platt R. Comparative effectiveness of different beta-adrenergic antagonists on mortality among adults with heart failure in clinical practice. *Arch Intern Med* 2008; **168**: 2415–21.
10. Heckbert SR, Wiggins KL, Glazer NL, Dublin S, Psaty BM, Smith NL, *et al*. Antihypertensive treatment with ACE inhibitors or beta-blockers and risk of incident atrial fibrillation in a general hypertensive population. *Am J Hypertens* 2009; **22**: 538–44.
11. Go AS, Magid DJ, Wells B, Sung SH, Cassidy-Bushrow AE, Greenlee RT, *et al*. The Cardiovascular

Research Network: a new paradigm for cardiovascular quality and outcomes research. *Circ Cardiovasc Qual Outcomes* 2008; **1**: 138–47.
12. Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. *Med Care* 1995; **33**: 783–95.
13. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992; **45**: 613–9.
14. Putnam KG, Buist DS, Fishman P, Andrade SE, Boles M, Chase GA, *et al*. Chronic disease score as a predictor of hospitalization. *Epidemiology* 2002; **13**: 340–6.
15. Rassen JA, Solomon DH, Curtis JR, Herrinton L, Schneeweiss S. Privacy-maintaining propensity score-based pooling of multiple databases applied to a study of biologics. *Med Care* 2010; **48** (6 Suppl): S83–9.
16. 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**: 512–22.
17. Arterburn DE, Alexander GL, Calvi J, Coleman LA, Gillman MW, Novotny R, *et al*. Body mass index measurement and obesity prevalence in ten U.S. health plans. *Clin Med Res* 2010; **8**: 126–30.
18. Davis LS, Collen MF, Rubin L, Van Brunt EE. Computer-stored medical record. *Comput Biomed Res* 1968; **1**: 452–69.
19. Friedman GD, Collen MF. A method of monitoring adverse drug reactions. In: LeCam LM, Neyman J, Scott EL, eds. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1972, VI, p. 367–82.
20. Friedman GD, Ury HK. Initial screening for carcinogenicity of commonly used drugs. *J Natl Cancer Inst* 1980; **65**: 723–33.
21. Friedman GD, Ury HK. Screening for possible drug carcinogenicity: second report of findings. *J Natl Cancer Inst* 1983; **71**: 1165–75.
22. Selby JV, Friedman GD, Fireman BH. Screening prescription drugs for possible carcinogenicity: eleven to fifteen years of follow-up. *Cancer Res* 1989; **49**: 5736–47.
23. Van Den Eeden SK, Friedman GD. Prescription drug screening for subsequent carcinogenicity. *Pharmacoepidemiol Drug Saf* 1995; **4**: 275–87.
24. Friedman GD. Barbiturates and lung cancer in humans. *J Natl Cancer Inst* 1981; **67**: 291–5.
25. Friedman GD, Habel LA. Barbiturates and lung cancer: a re-evaluation. *Int J Epidemiol* 1999; **28**: 375–9.
26. Friedman GD. Cancer after metronidazole. *N Engl J Med* 1980; **302**: 519.
27. Friedman GD. Rauwolfia and breast cancer: no relation found in long term users age fifty and over. *J Chronic Dis* 1983; **36**: 367–70.
28. Friedman GD. Digitalis and breast cancer. *Lancet* 1984; **2** (8407): 875.
29. Wagner EH, Greene SM, Hart G, Field TS, Fletcher S, Geiger AM, *et al*. Building a research consortium of large health systems: the Cancer Research Network. *J Natl Cancer Inst Monogr* 2005, pp. 3–11.
30. Hornbrook MC, Hart G, Ellis JL, Bachman DJ, Ansell G, Greene SM, *et al*. Building a virtual cancer research organization. *J Natl Cancer Inst Monogr* 2005, pp. 12–25.
31. 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** (6 Suppl): S45–51.
32. Platt R, Davis R, Finkelstein J, Go AS, Gurwitz JH, Roblin D, *et al*. Multicenter epidemiologic and health services research on therapeutics in the HMO Research Network Center for Education and Research on Therapeutics. *Pharmacoepidemiol Drug Saf* 2001; **10**: 373–7.
33. Donahue JG, Chan KA, Andrade SE, Beck A, Boles M, Buist DS, *et al*. Gastric and duodenal safety of daily alendronate. *Arch Intern Med* 2002; **162**: 936–42.
34. Brown JS, Kulldorff M, Chan KA, Davis RL, Graham D, Pettus PT, *et al*. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1275–84.
35. Brown JS, Kulldorff M, Petronis KR, Reynolds R, Chan KA, Davis RL, *et al*. Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations. *Pharmacoepidemiol Drug Saf* 2009; **18**: 226–34.
36. Smith DH, Kramer JM, Perrin N, Platt R, Roblin DW, Lane K, *et al*. A randomized trial of direct-to-patient communication to enhance adherence to beta-blocker therapy following myocardial infarction. *Arch Intern Med* 2008; **168**: 477–83; discussion 483; quiz 447.

37. Raebel MA, Carroll NM, Andrade SE, Chester EA, Lafata JE, Feldstein A, *et al*. Monitoring of drugs with a narrow therapeutic range in ambulatory care. *Am J Manag Care* 2006; **12**: 268–74.
38. Raebel MA, Carroll NM, Simon SR, Andrade SE, Feldstein AC, Lafata JE, *et al*. Liver and thyroid monitoring in ambulatory patients prescribed amiodarone in 10 HMOs. *J Manag Care Pharm* 2006; **12**: 656–64.
39. Raebel MA, McClure DL, Chan KA, Simon SR, Feldstein AC, Lafata JE, *et al*. Laboratory evaluation of potassium and creatinine among ambulatory patients prescribed spironolactone: are we monitoring for hyperkalemia? *Ann Pharmacother* 2007; **41**: 193–200.
40. Raebel MA, McClure DL, Simon SR, Chan KA, Feldstein A, Andrade SE, *et al*. Laboratory monitoring of potassium and creatinine in ambulatory patients receiving angiotensin converting enzyme inhibitors and angiotensin receptor blockers. *Pharmacoepidemiol Drug Saf* 2007; **16**: 55–64.
41. Raebel MA, McClure DL, Simon SR, Chan KA, Feldstein AC, Gunter MJ, *et al*. Frequency of serum creatinine monitoring during allopurinol therapy in ambulatory patients. *Ann Pharmacother* 2006; **40**: 386–91.
42. Raebel MA, Lyons EE, Andrade SE, Chan KA, Chester EA, Davis RL, *et al*. Laboratory monitoring of drugs at initiation of therapy in ambulatory care. *J Gen Intern Med* 2005; **20**: 1120–6.
43. Raebel MA, Chester EA, Newsom EE, Lyons EE, Kelleher JA, Long C, *et al*. Randomized trial to improve laboratory safety monitoring of ongoing drug therapy in ambulatory patients. *Pharmacotherapy* 2006; **26**: 619–26.
44. Raebel MA, Lyons EE, Chester EA, Bodily MA, Kelleher JA, Long CL, *et al*. Improving laboratory monitoring at initiation of drug therapy in ambulatory care: a randomized trial. *Arch Intern Med* 2005; **165**: 2395–401.
45. Smith DH, Feldstein AC, Perrin NA, Yang X, Rix MM, Raebel MA, *et al*. Improving laboratory monitoring of medications: an economic analysis alongside a clinical trial. *Am J Manag Care* 2009; **15**: 281–9.
46. Feldstein AC, Smith DH, Perrin N, Yang X, Rix M, Raebel MA, Magid DJ, *et al*. Improved therapeutic monitoring with several interventions: a randomized trial. *Arch Intern Med* 2006; **166**: 1848–54.
47. Palen TE, Raebel M, Lyons E, Magid DM. Evaluation of laboratory monitoring alerts within a computerized physician order entry system for medication orders. *Am J Manag Care* 2006; **12**: 389–95.
48. Raebel MA, Ross C, Cheetham C, Petersen H, Saylor G, Smith DH, *et al*. Increasingly restrictive definitions of hyperkalemia outcomes in a database study: effect on incidence estimates. *Pharmacoepidemiol Drug Saf* 2010; **19**: 19–25.
49. Raebel MA, Ross C, Xu S, Roblin DW, Cheetham C, Blanchette CM, *et al*. Diabetes and drug-associated hyperkalemia: effect of potassium monitoring. *J Gen Intern Med* 2010; **25**: 326–33.
50. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health* 2010; **13**: 273–7.
51. Feldstein A,Simon SR, Schneider J, Krall M, Laferriere D, Smith DH, *et al*. How to design computerized alerts to safe prescribing practices. *Jt Comm J Qual Saf* 2004; **30**: 602–13.
52. Feldstein A, *et al*. *Design and Implementation of Decision Support for Outpatient Prescribing in an Electronic Medical Record. The Safety in Prescribing Study*. Rockville, MD: Agency for Healthcare Research and Quality, Publication No. 050021, 2005.
53. Feldstein AC, Smith DH, Perrin N, Yang X, Simon SR, Krall M, *et al*. Reducing warfarin medication interactions: an interrupted time series evaluation. *Arch Intern Med* 2006; **166**: 1009–15.
54. Simon SR, Smith DH, Feldstein AC, Perrin N, Yang X, Zhou Y, *et al*. Computerized prescribing alerts and group academic detailing to reduce the use of potentially inappropriate medications in older people. *J Am Geriatr Soc* 2006; **54**: 963–8.
55. Smith DH, Perrin N, Feldstein A, Yang X, Kuang D, Simon SR, *et al*. The impact of prescribing safety alerts for elderly persons in an electronic medical record: an interrupted time series evaluation. *Arch Intern Med* 2006; **166**: 1098–104.
56. Feldstein A, Simon SR, Schneider J, Krall M, Laferriere D, Smith DH, *et al*. How to design computerized alerts to safe prescribing practices. *Jt Comm J Qual Saf* 2004; **30**: 602–13.
57. Buist DS, Newton KM, Miglioretti DL, Beverly K, Connelly MT, Andrade S, *et al*. Hormone therapy prescribing patterns in the United States. *Obstet Gynecol* 2004; **104**: 1042–50.
58. Newton KM, Buist DS, Yu O, Hartsfield CL, Andrade SE, Wei F, *et al*. Hormone therapy initiation after the Women's Health Initiative. *Menopause* 2008; **15**: 487–93.

59. Owusu C, Buist DS, Field TS, Lash TL, Thwin SS, Geiger AM, *et al*. Predictors of tamoxifen discontinuation among older women with estrogen receptor-positive breast cancer. *J Clin Oncol* 2008; **26**: 549–55.
60. Aiello EJ, Buist DS, Wagner EH, Tuzzio L, Greene SM, Lamerato LE, *et al*. Diffusion of aromatase inhibitors for breast cancer therapy between 1996 and 2003 in the Cancer Research Network. *Breast Cancer Res Treat* 2008; **107**: 397–403.
61. Buist DS, Chubak J, Prout M, Yood MU, Bosco JL, Thwin SS, *et al*. Referral, receipt, and completion of chemotherapy in patients with early-stage breast cancer older than 65 years and at high risk of breast cancer recurrence. *J Clin Oncol* 2009; **27**: 4508–14.
62. Enger SM, Thwin SS, Buist DS, Field T, Frost F, Geiger AM, *et al*. Breast cancer treatment of older women in integrated health care settings. *J Clin Oncol* 2006; **24**: 4377–83.
63. Asgari MM, Tang J, Epstein EH Jr, Chren MM, Warton EM, Quesenberry CP, *et al*. Statin use and risk of basal cell carcinoma. *J Am Acad Dermatol* 2009; **61**: 66–72.
64. Boudreau DM, Koehler E, Rulyak SJ, Haneuse S, Harrison R, Mandelson MT, *et al*. Cardiovascular medication use and risk for colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 3076–80.
65. Boudreau DM, Yu O, Miglioretti DL, Buist DS, Heckbert SR, Daling JR. Statin use and breast cancer risk in a large population-based setting. *Cancer Epidemiol Biomarkers Prev* 2007; **16**: 416–21.
66. Yu O, Boudreau DM, Buist DS, Miglioretti DL. Statin use and female reproductive organ cancer risk in a large population-based setting. *Cancer Causes Control* 2009; **20**: 609–16.
67. Fagerlin A, Zikmund-Fisher BJ, Nair V, Derry HA, McClure JB, Greene S, *et al*. Women's decisions regarding tamoxifen for breast cancer prevention: responses to a tailored decision aid. *Breast Cancer Res Treat* 2010; **119**: 613–20.
68. Malin JL, Ko C, Ayanian JZ, Harrington D, Nerenz DR, Kahn KL, *et al*. Understanding cancer patients' experience and outcomes: development and pilot study of the Cancer Care Outcomes Research and Surveillance patient survey. *Support Care Cancer* 2006; **14**: 837–48.
69. Platt R, Takvorian SU, Septimus E, Hickok J, Moody J, Perlin J, *et al*. Cluster randomized trials in comparative effectiveness research: randomizing hospitals to test methods for prevention of healthcare-associated infections. *Med Care* 2010; **48** (6 Suppl): S52–7.
70. Davis RL, Kolczak M, Lewis E, Nordin J, Goodman M, Shay DK, *et al*. Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology* 2005; **16**: 336–41.
71. Ritzwoller D, Bridges CB, Shetterly S, Yamasaki K, Kolczak M, France EK. Effectiveness of the 2003–2004 influenza vaccine among children 6 months to 8 years of age, with 1 vs 2 doses. *Pediatrics* 2005; **116**: 153–9.
72. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *Int J Epidemiol* 2006; **35**: 337–44.
73. Jackson LA, Nelson JC, Benson P, Neuzil KM, Reid RJ, Psaty BM, *et al*. Functional status is a confounder of the association of influenza vaccine and risk of all cause mortality in seniors. *Int J Epidemiol* 2006; **35**: 345–52.
74. Jackson ML, Nelson JC, Benson P, Neuzil KM, Reid RJ, Psaty BM, *et al*. To rule out confounding, observational studies of influenza vaccine need to include analyses during the "preinfluenza period". *Arch Intern Med* 2007; **167**: 1553–4; author reply 1554–5.
75. Glanz JM, McClure DL, Magid DJ, Daley MF, France EK, Salmon DA, *et al*. Parental refusal of pertussis vaccination is associated with an increased risk of pertussis infection in children. *Pediatrics* 2009; **123**: 1446–51.
76. Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K, *et al*. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 2007; **45** (10 Supl 2): S89–95.
77. Klein N, Fireman B, Yih WK, Lewis E, Kulldorff M, Ray P, *et al*. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. *Pediatrics* 2010; **126**: e1–8.
78. Ho M, Zeng C, Tavel HM, Selby JV, O'Connor PJ, Margolis KL, *et al*. Trends in first-line therapy for hypertension in the Cardiovascular Research Network Hypertension Registry, 2002–2007. *Arch Intern Med* 2010; **170**: 912–3.
79. Selby JV, Lee J, Swain BE, Tavel HM, Ho PM, Margolis KL, *et al*. Trends in time to confirmation and recognition of new-onset hypertension, 2002–2006. *Hypertension* 2010; **56**: 605–11.
80. Magid DJ, Shetterly SM, Margolis KL, Tavel HM, O'Connor PJ, Selby JV, *et al*. Comparative effectiveness of angiotensin-converting enzyme inhibitors

versus beta-blockers as second-line therapy for hypertension. *Circ Cardiovasc Qual Outcomes* 2010; **3**: 453–8.

81. Andrade SE, Gurwitz JH, Davis RL, Chan KA, Finkelstein JA, Fortman K, *et al*. Prescription drug use in pregnancy. *Am J Obstet Gynecol* 2004; **191**: 398–407.
82. Andrade SE, McPhillips H, Loren D, Raebel MA, Lane K, Livingston J, *et al*. Antidepressant medication use and risk of persistent pulmonary hypertension of the newborn. *Pharmacoepidemiol Drug Saf* 2009; **18**: 246–52.
83. Andrade SE, Raebel MA, Brown J, Lane K, Livingston J, Boudreau D, *et al*. Outpatient use of cardiovascular drugs during pregnancy. *Pharmacoepidemiol Drug Saf* 2008; **17**: 240–7.
84. Andrade SE, Raebel MA, Brown J, Lane K, Livingston J, Boudreau D, *et al*. Use of antidepressant medications during pregnancy: a multisite study. *Am J Obstet Gynecol* 2008; **198**: 194 e1–5.
85. Andrade SE, Raebel MA, Morse AN, Davis RL, Chan KA, Finkelstein JA, *et al*. Use of prescription medications with a potential for fetal harm among pregnant women. *Pharmacoepidemiol Drug Saf* 2006; **15**: 546–54.
86. Davis RL, Rubanowice D, McPhillips H, Raebel MA, Andrade SE, Smith D, *et al*. Risks of congenital malformations and perinatal events among infants exposed to antidepressant medications during pregnancy. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1086–94.
87. Davis RL, Andrade S, Rubanowice D, McPhillips H, Boudreau D, Raebel M, *et al*. Risks to the newborn associated with in-utero exposure to beta-blockers and calcium-channel blockers. *Clin Med Res* 2010; **8**: 57 (Abstract C-A4–03).
88. Graham DJ, Staffa JA, Shatin D, Andrade SE, Schech SD, La Grenade L, *et al*. Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs. *JAMA* 2004; **292**: 2585–90.
89. Smalley W, Shatin D, Wysowski DK, Gurwitz J, Andrade SE, Goodman M, *et al*. Contraindicated use of cisapride: impact of food and drug administration regulatory action. *JAMA* 2000; **284**: 3036–9.
90. Bayliss EA, Ellis JL, Delate T, Steiner JF, Raebel MA, *et al*. Characteristics of Medicare Part D beneficiaries who reach the drug benefit threshold in both of the first two years of the Part D benefit. *Med Care* 2010; **48**: 267–72.
91. Harrold LR, Andrade SE, Briesacher B, Raebel MA, Fouayzi H, Yood RA, *et al*. The dynamics of chronic gout treatment: medication gaps and return to therapy. *Am J Med* 2010; **123**: 54–9.
92. Harrold LR, Andrade SE, Briesacher BA, Raebel MA, Fouayzi H, Yood RA, *et al*. Adherence with urate-lowering therapies for the treatment of gout. *Arthritis Res Ther* 2009; **11**: R46.
93. Raebel MA, Delate T, Ellis JL, Bayliss EA. Effects of reaching the drug benefit threshold on Medicare members' healthcare utilization during the first year of Medicare Part D. *Med Care* 2008; **46**: 1116–22.
94. Farber HJ, Capra AM, Finkelstein JA, Lozano P, Quesenberry CP, Jensvold NG, *et al*. Misunderstanding of asthma controller medications: association with nonadherence. *J Asthma* 2003; **40**: 17–25.
95. Farber HJ, Chi FW, Capra A, Jensvold NG, Finkelstein JA, Lozano P, *et al*. Use of asthma medication dispensing patterns to predict risk of adverse health outcomes: a study of Medicaid-insured children in managed care programs. *Ann Allergy Asthma Immunol* 2004; **92**: 319–28.
96. Boudreau DM, Malone DC, Raebel MA, Fishman PA, Nichols GA, Feldstein AC, *et al*. Health care utilization and costs by metabolic syndrome risk factors. *Metab Syndr Relat Disord* 2009; **7**: 305–14.
97. Levin TR, Schmittdiel JA, Henning JM, Kunz K, Henke CJ, Colby CJ, *et al*. A cost analysis of a Helicobacter pylori eradication strategy in a large health maintenance organization. *Am J Gastroenterol* 1998; **93**: 743–7.
98. Ray GT, Butler JC, Black SB, Shinefield HR, Fireman BH, Lieu TA. Observed costs and health care use of children in a randomized controlled trial of pneumococcal conjugate vaccine. *Pediatr Infect Dis J* 2002; **21**: 361–5.
99. Johnson RE, McFarland BH. Antipsychotic drug exposure in a health maintenance organization. *Med Care* 1993; **31**: 432–44.
100. Johnson RE, McFarland BH, Nichols GA. Changing patterns of antidepressant use and costs in a health maintenance organisation. *Pharmacoeconomics* 1997; **11**: 274–86.
101. Johnson RE, McFarland BH. Lithium use and discontinuation in a health maintenance organization. *Am J Psychiatry* 1996; **153**: 993–1000.
102. Johnson RE, McFarland BH, Charlotte A, Corelle CA, Gary T, Woodson GT, *et al*. Estimating daily dose for pharmacoepidemiologic studies: alprazolam as an example. *Pharmacoepidemiol Drug Saf* 1994; **3**: 139–45.
103. Johnson RE, McFarland B, Woodson GT. Whither triazolam? *Med Care* 1997; **35**: 303–10.

104. McFarland BH. Comparing period prevalences with application to drug utilization. *J Clin Epidemiol* 1996; **49**: 473–82.
105. Keith DS, Nichols GA, Gullion CM, Brown JB, Smith DH. Longitudinal follow-up and outcomes among a population with chronic kidney disease in a large managed care organization. *Arch Intern Med* 2004; **164**: 659–63.
106. Smith DH, Gullion CM, Nichols G, Keith DS, Brown JB. Cost of medical care for chronic kidney disease and comorbidity among enrollees in a large HMO population. *J Am Soc Nephrol* 2004; **15**: 1300–6.
107. Goodwin FK, Fireman B, Simon GE, Hunkeler EM, Lee J, Revicki D. Suicide risk in bipolar disorder during treatment with lithium and divalproex. *JAMA* 2003; **290**: 1467–73.
108. McCarty CA, Burmester JK, Mukesh BN, Patchett RB, Wilke RA. Intraocular pressure response to topical beta-blockers associated with an ADRB2 single-nucleotide polymorphism. *Arch Ophthalmol* 2008; **126**: 959–63.
109. McCarty CA, Chapman-Stone D, Derfus T, Giampietro PF, Fost N; Marshfield Clinic PMRP Community Advisory Group. Community consultation and communication for a population-based DNA biobank: the Marshfield clinic personalized medicine research project. *Am J Med Genet A* 2008; **146A**: 3026–33.
110. Onitilo AA, McCarty CA, Wilke RA, Glurich I, Engel JM, Flockhart DA, *et al*. Estrogen receptor genotype is associated with risk of venous thromboembolism during tamoxifen therapy. *Breast Cancer Res Treat* 2009; **115**: 643–50.
111. Peissig P, Sirohi E, Berg RL, Brown-Switzer C, Ghebranious N, McCarty CA, *et al*. Construction of atorvastatin dose-response relationships using data from a large population-based DNA biobank. *Basic Clin Pharmacol Toxicol* 2007; **100**: 286–8.
112. Wilke RA, Berg RL, Linneman JG, Zhao C, McCarty CA, Krauss RM. Characterization of low-density lipoprotein cholesterol-lowering efficacy for atorvastatin in a population-based DNA biorepository. *Basic Clin Pharmacol Toxicol* 2008; **103**: 354–9.
113. Fortuny J, Johnson CC, Bohlke K, Chow WH, Hart G, Kucera G, *et al*. Use of anti-inflammatory drugs and lower esophageal sphincter-relaxing drugs and risk of esophageal and gastric cancers. *Clin Gastroenterol Hepatol* 2007; **5**: 1154–93.
114. Giampietro PF, McCarty C, Mukesh B, McKiernan F, Wilson D, Shuldiner A, *et al*. The role of cigarette smoking and statins in the development of postmenopausal osteoporosis: a pilot study utilizing the Marshfield Clinic Personalized Medicine Cohort. *Osteoporos Int* 2010; **21**: 467–77.
115. Harrold LR, Andrade SE. Medication adherence of patients with selected rheumatic conditions: a systematic review of the literature. *Semin Arthritis Rheum* 2009; **38**: 396–402.
116. Aiello Bowles EJ, Tuzzio L, Ritzwoller DP, Williams AE, Ross T, Wagner EH, *et al*. Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. *Med Care* 2009; **47**: 1091–7.
117. Boudreau DM, Buist DS, Rutter CM, Fishman PA, Beverly KR, Taplin S. Impact of hormone therapy on false-positive recall and costs among women undergoing screening mammography. *Med Care* 2006; **44**: 62–9.
118. Ferrara A, Quesenberry CP, Karter AJ, Njoroge CW, Jacobson AS, Selby JV; Northern California Kaiser Permanente Diabetes Registry. Current use of unopposed estrogen and estrogen plus progestin and the risk of acute myocardial infarction among women with diabetes: the Northern California Kaiser Permanente Diabetes Registry, 1995–1998. *Circulation* 2003; **107**: 43–8.
119. Bambauer KZ, Soumerai SB, Adams AS, Mah C, Zhang F, McLaughlin TJ. Does antidepressant adherence have an effect on glycemic control among diabetic antidepressant users? *Int J Psychiatry Med* 2004; **34**: 291–304.
120. Lo Sasso AT, Rost K, Beck A. Modeling the impact of enhanced depression treatment on workplace functioning and costs: a cost–benefit approach. *Med Care* 2006; **44**: 352–8.
121. Simon GE, Katon WJ, Lin EH, Rutter C, Manning WG, Von Korff M, *et al*. Cost-effectiveness of systematic depression treatment among people with diabetes mellitus. *Arch Gen Psychiatry* 2007; **64**: 65–72.
122. Simon GE, Savarino J. Suicide attempts among patients starting depression treatment with medications or psychotherapy. *Am J Psychiatry* 2007; **164**: 1029–34.
123. Soumerai SB, McLaughlin TJ, Ross-Degnan D, Casteris CS, Bollini P. Effects of a limit on 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**: 650–5.

124. Chan KA, Truman A, Gurwitz JH, Hurley JS, Martinson B, Platt R, *et al*. A cohort study of the incidence of serious acute liver injury in diabetic patients treated with hypoglycemic agents. *Arch Intern Med* 2003; **163**: 728–34.
125. Nichols GA, Conner C, Brown JB. Initial nonadherence, primary failure and therapeutic success of metformin monotherapy in clinical practice. *Curr Med Res Opin* 2010; **26**: 2127–35.
126. O'Connor PJ. Improving medication adherence: challenges for physicians, payers, and policy makers. *Arch Intern Med* 2006; **166**: 1802–4.
127. Selby JV, Uratsu CS, Fireman B, Schmittdiel JA, Peng T, Rodondi N, *et al*. Treatment intensification and risk factor control: toward more clinically relevant quality measures. *Med Care* 2009; **47**: 395–402.
128. Dublin S, Jackson ML, Nelson JC, Weiss NS, Larson EB, Jackson LA. Statin use and risk of community acquired pneumonia in older people: population based case-control study. *BMJ* 2009; **338**: b2137.
129. Greene SK, Kulldorff M, Lewis EM, Li R, Yin R, Weintraub ES, *et al*. Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project. *Am J Epidemiol* 2010; **171**: 177–88.
130. Jackson LA, Yu O, Nelson J, Belongia EA, Hambidge SJ, Baxter R, *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**: 4912–6.
131. Nelson JC, Bittner RC, Bounds L, Zhao S, Baggs J, Donahue JG, *et al*. Compliance with multiple-dose vaccine schedules among older children, adolescents, and adults: results from a vaccine safety datalink study. *Am J Public Health* 2009; **99** (Suppl 2): S389–97.
132. Stille CJ, Andrade SE, Huang SS, Nordin J, Raebel MA, Go AS, *et al*. Increased use of second-generation macrolide antibiotics for children in nine health plans in the United States. *Pediatrics* 2004; **114**: 1206–11.
133. Adams RJ, Fuhlbrigge AL, Finkelstein JA, Weiss ST. Intranasal steroids and the risk of emergency department visits for asthma. *J Allergy Clin Immunol* 2002; **109**: 636–42.
134. Finkelstein JA, Lozano P, Streiff KA, Arduino KE, Sisk CA, Wagner EH, *et al*. Clinical effectiveness research in managed-care systems: lessons from the Pediatric Asthma Care PORT. Patient Outcomes Research Team. *Health Serv Res* 2002; **37**: 775–89.
135. Harrold LR, Patterson MK, Andrade SE, Dube T, Go AS, Buist AS, *et al*. Asthma drug use and the development of Churg-Strauss syndrome (CSS). *Pharmacoepidemiol Drug Saf* 2007; **16**: 620–6.
136. Boudreau D, Von Korff M, Rutter CM, Saunders K, Ray GT, Sullivan MD, *et al*. Trends in long-term opioid therapy for chronic non-cancer pain. *Pharmacoepidemiol Drug Saf* 2009; **18**: 1166–75.
137. Dunn KM, Saunders KW, Rutter CM, Banta-Green CJ, Merrill JO, Sullivan MD, *et al*. Opioid prescriptions for chronic pain and overdose: a cohort study. *Ann Intern Med* ; **152**: 85–92.
138. Korff MV, Saunders K, Thomas Ray G, Boudreau D, Campbell C, Merrill J, *et al*. De facto long-term opioid therapy for noncancer pain. *Clin J Pain* 2008; **24**: 521–7.
139. Rolnick SJ, Jackson J, Nelson WW, Butani A, Herrinton LJ, Hornbrook M, *et al*. Pain management in the last six months of life among women who died of ovarian cancer. *J Pain Symptom Manage* 2007; **33**: 24–31.
140. Weisner CM, Campbell CI, Ray GT, Saunders K, Merrill JO, Banta-Green C, *et al*. Trends in prescribed opioid therapy for non-cancer pain for individuals with prior substance use disorders. *Pain* 2009; **145**: 287–93.
141. Raebel MA, Ellis JL, Andrade SE. Evaluation of gestational age and admission date assumptions used to determine prenatal drug exposure from administrative data. *Pharmacoepidemiol Drug Saf* 2005; **14**: 829–36.
142. Brown JS, Moore KM, Braun MM, Ziyadeh N, Chan KA, Lee GM, *et al*. Active influenza vaccine safety surveillance: potential within a healthcare claims environment. *Med Care* 2009; **47**: 1251–7.
143. Glasgow RE, Magid DJ, Beck A, Ritzwoller D, Estabrooks PA, *et al*. Practical clinical trials for translating research to practice: design and measurement recommendations. *Med Care* 2005; **43**: 551–7.
144. Hubbard RA, Miglioretti DL, Smith RA. Modelling the cumulative risk of a false-positive screening test. *Stat Methods Med Res* 2010; **19**: 429–49.
145. Moore KM, Duddy A, Braun MM, Platt R, Brown JS, *et al*. Potential population-based electronic data sources for rapid pandemic influenza vaccine adverse event detection: a survey of health plans. *Pharmacoepidemiol Drug Saf* 2008; **17**: 1137–41.
146. Lafata JE, Gunter MJ, Hsu J, Kaatz S, Krajenta R, Platt R, *et al*. Academic detailing to improve laboratory testing among outpatient medication users. *Med Care* 2007; **45**: 966–72.

147. Lafata JE, Schultz L, Simpkins J, Chan KA, Horn JR, Kaatz S, *et al*. Potential drug-drug interactions in the outpatient setting. *Med Care* 2006; **44**: 534–41.
148. Raebel MA, Carroll NM, Kelleher JA, Chester EA, Berga S, Magid DJ. Randomized trial to improve prescribing safety during pregnancy. *J Am Med Inform Assoc* 2007; **14**: 440–50.
149. Raebel MA, Charles J, Dugan J, Carroll NM, Korner EJ, Brand DW, *et al*. Randomized trial to improve prescribing safety in ambulatory elderly patients. *J Am Geriatr Soc* 2007; **55**: 977–85.

CHAPTER 13
Commercial Insurance Databases

John Seeger[1] *and Gregory W. Daniel*[2]

[1] Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[2] HealthCore, Inc. Wilmington, DE, USA

Introduction

A large, representative cohort of people for whom virtually all health-care interactions are captured automatically and from which study-specific data could readily be pulled and linked to other data sources can be an ideal setting within which to conduct pharmacoepidemiologic research. Within such databases, information on enrollment status, medication exposure and use patterns, medical services, and diagnoses for eligible members is captured as a routine part of reimbursing health care. These transactional records are maintained in a searchable form and can be integrated to create longitudinal claims-based histories for individuals represented in the data. In addition to capturing information on all reimbursed services across the health-care setting, regardless of provider or facility, health insurance claims databases can also be linked with supplemental data, such as patients' medical records, disease registries, and patient and/or provider surveys, in order to capture a more complete picture of health-care experiences. However, several important aspects of commercial insurance databases influence both the conduct of the research and the interpretation of its results. This chapter will explore the advantages and research potential offered by commercial databases for pharmacoepidemiology while also noting some of the challenges to their use.

Databases derived from commercial health insurers typically consist of current and historical medical and pharmacy claims for reimbursement of health-care services regardless of setting, along with enrollment information on the insured population.[1] Therefore, these databases represent longitudinal collections of billable health-care interactions. The geographic representation of the database might vary from being local or regional to being nationwide, and the insured people may be covered by a variety of insurance products, so that some people have both medical and pharmacy coverage, while other aspects such as mental health, vision, and dental services are included for some. Analyses can be restricted to subsets of a particular database, such as those members having both medical and pharmacy benefits.

The diversity of the underlying health insurance benefit plans within these databases is a source of variation in utilization of health services that may not be apparent. Each of the individual health plans within a larger commercial health insurance database may offer variations on a basic benefit design which is customized to the needs and budget of the organizations or individuals who are covered by the insurance. These differences in benefit designs may manifest as differences in incentives to the members (patients) and to the providers to utilize health-care services. These incentives may contribute to variations in health-care utilization

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

that reflect the incentives rather than patients' clinical status. Some of the incentives are readily ascertained, such as differences in co-payment and co-insurance amounts, but other incentives may be less apparent to a researcher using a claims database, such as formulary restrictions, stepped therapy requirements, member participation in disease or care management programs, and variations in provider reimbursement. In addition to variation across plans with respect to incentives, coding for individual data elements may not be uniform across the health plans. Further, particular benefits such as mental health or other specialty services may be "carved out" of the benefit and thus may not be captured by the health plan for that particular account. Also, patients with certain forms of Medicare (e.g., Medicare supplemental plans) may be included within commercial insurance databases, and the nature of the benefits may vary so that the database files do not capture all claims for such enrollees.

Some of the information on health insurance coverage may be considered confidential and proprietary by the health plan (e.g., benefit design information), or may require access to individually identifiable information in a way that is not consistent with adherence to privacy policies. As a result, details or data that might reveal variation with respect to some of the benefit structures may not be available in externally licensed versions of the data from commercial health insurers. In such settings, it is advantageous to work directly with the health plan, such as through researchers employed by the health plan who are experienced in the use of the data for research.

When used appropriately, commercial health insurer data can provide important information about the safety and effectiveness of drugs and other medical interventions derived from patients receiving those interventions in actual practice. However, the conduct and interpretation of studies based on commercial databases can be subject to numerous limitations, including bias that might derive from a lack of correspondence between individual insurance claims and a patient's underlying conditions.[2] These limitations may not be apparent to researchers unless they have experience conducting research within such databases.

The data incorporated in a health insurance claim reflect on the patient for whom the claim is submitted, but not necessarily in a straightforward way. Considering the nature of the process that gave rise to the insurance claim allows the researcher to understand the data better. Health insurance claims data derive from a system that is primarily an accounting tool for tracking reimbursement for health-care expenses related to medical services provided to eligible members of the health plan. Therefore, health insurance claims data arise from a system not designed for research or clinical care. Thus, research using such data needs to accommodate the fact that insurance claims represent financial transactions related to patients' health-care utilization rather than a clinical record of the health care. It is for this reason that linking the insurance claims data with supplemental data gathered from providers (e.g., patients' medical charts, electronic laboratory results, provider surveys, etc.), disease registries, or the patients themselves via survey-based questionnaires can be advantageous for validating characteristics or events identified from insurance claims and for supplementing the claims with additional variables that cannot be obtained from claims alone.

Valid interpretation of data within a commercial health insurance claims database depends on understanding how health-care interactions are reflected as individual claims for reimbursement. Individual claims are valid in the sense that they are actual claims for reimbursement, but the interpretation of the claim as clinical information can be invalid. Accordingly, claims are suggestive, but not definitive, so that some capacity to go beyond the information contained in the database improves the validity of inferences made about populations represented within the database.

Careful consideration of the systems that give rise to health insurance claims and their mechanism of incorporation into a research database is a necessary but insufficient element of high quality research conducted within such a database. Although many pharmacoepidemiologic research

questions can be adequately answered using the health insurance data alone, it is often necessary to go beyond the health insurance claims data. At one extreme, the limitations of the health insurance database may be considered insurmountable so that the idea of using the database would be discarded altogether in favor of conducting the research through a clinical trial (see Chapter 36), *ad hoc* non-randomized study (Chapter 22), or registry (Chapter 21). Often, however, a hybrid approach can be used that involves using both the health insurance database and either linkage to another data source or primary data collection nested within the health insurer. This approach retains the strengths of the commercial health insurance database in terms of its breadth, its recording of routine care, and its ability to rapidly assemble cohorts and identify outcomes among them, but it expands on the database by adding depth to the data, allowing investigators to enrich health insurance claims data with other data elements that might be needed for a particular research question. The range of potential additions is essentially limitless, and may include laboratory results, information contained in medical records, information gathered within disease, immunization, or national vital registries, or information obtained through direct contact with physicians or patients. When accomplished, this data expansion enables pharmacoepidemiologic research to be conducted within a 'data environment' rather than within a single database. Conducting research within such a data environment can address many of the limitations inherent in insurance claims databases, such as lack of data on important potential confounders or patient subgroups, or uncertainty about whether a claim suggesting an outcome actually means the patient experienced the outcome. Further, such research likely involves working closely with the health plan to obtain necessary approvals, and strict maintenance of confidentiality. This process is time-consuming, but ensures data integrity, confidentiality, respects the primacy of the doctor–patient relationship, and seeks to minimize provider burden. Depending on the source of data beyond claims, professionals with a variety of clinical experience may be employed by the health plan or hired as independent contractors to facilitate data collection.

Numerous commercial health insurance databases have been used for pharmacoepidemiologic research, and the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) lists 193 databases worldwide (91 in the US). (available at: http://www.ispor.org/digestofintdb/countrylist.aspx) Not all of these databases are commercial health insurance databases, but many of them share features with commercial health insurance databases so that considerations for their use and suitability for answering pharmacoepidemiologic research questions is similar.

Description

Health insurance claims databases used for research typically comprise linkable files, each with listings of discrete health insurance claims. These separate files represent different data sources (e.g., medical claims, pharmacy claims, and enrollment files) that can be compiled for research, for example by linking together claims corresponding to individual patients in a relational database.[3] This linkage might be accomplished through a unique patient identifier. Such databases that are available commercially might employ any of a range of de-identification mechanisms in order to comply with confidentiality provisions, and these may need to be taken into account when conducting research. However, caution should be noted that, if using a de-identified commercial insurance claims database, it is virtually impossible to link such data with external data described in the section above by virtue of the fact that all personal identifiers will have been eliminated. Typical commercial database files and the data elements contained within are shown in Table 13.1.

The flexibility of these databases derives from their origin as a means to track health-care services for reimbursement among people whose insurance covers a broad range of medical services. The resulting collection of transactional data between providers and the health insurer covers virtually all billable health-care services, permitting a wide

Table 13.1. Typical commercial database file structure and the data elements within it

File	Data source	Expected content
Membership	Application forms and renewals	Data elements include dates of enrollment and disenrollment, basic demographic information such as patient identifier number, date of birth, gender, and zip code, along with benefit package.
Medical claims	Physician offices and health-care institutions	Claims submitted for reimbursement of health-care services provided are retained in this file. Medical claims derive from various sites where the services are provided (e.g., inpatient, hospital outpatient, emergency room, surgery center, physician's office) for virtually all types of covered services, including specialty, preventive, and office-based injections, laboratory services and diagnostic imaging, and other treatment. Claims are submitted electronically or by mail. Claims are submitted in UB-92 or HCFA 1500 formats and included on the claim are patient and provider identifiers, dates of service, and details of the service coded by procedure codes (CPT, HCPCS, and ICD-9) and diagnosis codes (by ICD-9).
Pharmacy claims	Pharmacies	Claims for medications dispensed to insured people are usually submitted electronically by the pharmacy at the time a prescription is filled. Each claim specifies what was dispensed and when, including the drug name, date dispensed, dosage of medication dispensed, duration of the prescription in days, and quantity dispensed. Data on the use of inpatient drugs are not generally available.
Professional data	Provider database	A separate file contains data on the health plan's participating physicians and other health professionals, including type and location, as well as physician specialty or subspecialty. A unique identification number is assigned to each health professional and institution. Precautions are taken to protect the identity of health professionals.

range of pharmacoepidemiologic questions to be addressed.[4]

Health insurance claims are typically submitted for reimbursement electronically by health-care providers. Through the electronic claims adjudication process, claims transaction systems complete autoadjudication to determine the majority of the claims' status. Claims adjudication is the processing of a claim through a series of edits to determine proper payment. Claims may be approved, with payment levels depending upon benefit design, status of deductible or out-of-pocket caps, etc., denied with no payment, or pended, which may require more information or review before approval or denial can be determined. Approved payment levels can range from being 100% covered by the health plan up to the allowed amount for the service to 100% member responsibility (e.g., if the member has a deductible and the cost of the service is less than the remaining annual deductible amount). Types of claims include non-capitated outpatient, emergency department, and inpatient encounters, physicians' services, durable medical equipment, ancillary services (including claims for laboratory tests), as well as outpatient pharmacy dispensing for health-plan members with eligibility at the time of service.

In the course of conducting research involving data from a commercial health insurer, a number of activities are common:

• Defining disease, exposure, and confounding variables. This is done by creating user code lists through appropriate aggregates of diagnosis or procedure codes for disease variables, and specific

codes for drug exposures. These code lists identify cohorts of patients from the database and are used for exposure, covariate, and outcome ascertainment.

• Linking records and files. Enrollment, medical claims, pharmacy claims, and physician claims must be combined by linking members' discrete records, inpatient claims, outpatient claims, and pharmacy claims to member and health-professional data. These linkages allow for the creation of study-specific analytic files (e.g., flat files containing one record per subject meeting inclusion criteria) to be derived from the relational database represented by the insurance enrollment and claims data. This can include linking across people, such as linking mothers with infants for studies of the effect of drug exposure during pregnancy.

• Constructing longitudinal histories. Information on diagnosis, treatments, and the occurrence of adverse clinical events, as coded on claims, can be tracked across time for individual people or for groups of them. These data can be used to evaluate adherence to recommended monitoring guidelines or persistence of use for specific medications (see also Chapter 42).

• Identifying denominators. All members eligible for health-care reimbursement and dates of eligibility can be identified within commercial databases, so that the source population within which claims arise is fully enumerated and this can be used to determine population-based prevalence or incidence measures. Since dates of entry to and exit from the cohort are known (dates of eligibility for reimbursement) for each member, the effect on the at-risk person-time of partial-year enrollment can be handled analytically. The populations can be narrowed by age, gender, benefit type, time period, duration of enrollment, or geography, so that prevalence or incidence measures can be defined within any of these subgroups.

• Identifying treatment at a particular point in time. Claims for drugs, whether dispensed to a patient by a pharmacy or administered to a patient by a health-care professional in the outpatient (but generally not inpatient) setting can be identified, as well as non-drug treatments. This provides the ability to identify and track specific treatments to specific patients and the timing and setting of the treatments.

• Provider characteristics. Numerous characteristics of health-care providers can be identified. This information can be used to locate a medical record for detailed clinical information not captured in the claims data or be used for research related to variations in practice patterns.

• Identifying drug exposure periods and person-time at risk of event occurrence. Health insurance databases contain the date on which prescriptions are filled, the amount dispensed, the duration of the prescription (days' supply), and the end of enrollment for each member. The dosage form, strength, and amount dispensed, along with days' supply fields can be used to estimate doses received at specific times and cumulative dose received.

• Enrichment of claims data when possible or where available. When using claims data within a larger data environment, enrichment with paper or electronic medical records, data obtained directly from patients or their physicians through surveys, or data routinely collected in disease, immunization, or national vital registries may be retrievable and linkable with the claims data.

Performing the activities described here requires extensive programming skills and data management and processing skills as well as extensive quality control checks to create the data sets suitable to address a particular research question.

Strengths

Commercial health insurance claims databases and data environments have strengths for pharmacoepidemiologic research related to their large size and their ability to record virtually any billable health-care interaction. The health insurer has an incentive to maintain accurate records of when a person is eligible to receive reimbursement for health-care services and when they are no longer eligible for reimbursement. This accurate recording of enrollment translates directly into an enumeration of the open cohort from whom the health insurance claims transactions will arise. The availability of enrollment data with unique health-plan

identification numbers provides several key strengths to using health insurance claims data, as they enable investigators to identify persons at risk, making for straightforward estimation of incidence and prevalence. In addition, the enrollment data enables the investigator to identify periods of time over which persons are most likely under observation and less likely to seek care for covered services outside of the plan.

Physicians' office staff have incentives to submit claims for reimbursement of services performed rapidly and accurately in order to shorten the time to reimbursement. This incentive structure and the way health insurance billing is integrated into standard health-care workflow means that the record of health-care interactions is not influenced by any incremental burden that might be imposed by prospective data capture, as is generally the case in clinical trials, registries, or other designs such as the need to complete additional case report forms or the need to schedule follow-up visits by patients in order to capture additional information. This largely automatic data capture for purposes of reimbursement reduces the likelihood that the data capture itself might bias assessments of hypotheses. In addition, enrichment of claims data to create a data environment typically relies on existing data sources rather than prospectively captured data, although the latter is possible if the research needs require such form of data capture. The net result is a relatively short time lag from the date of service until the claim is incorporated into a usable research format. This incentive structure might lead to billing and coding practices that result in the highest payment regardless of the actual diagnosis, while periodic audits of billing provide a countering force that keeps such "upcoding" in check. However, the potential for mismatch between what is coded on a claim and what clinical condition a patient has should always be considered.

When conducting research within the context of the claims database or data environment, interpretation of findings must occur with an understanding of the purpose for which the data was collected. The data may faithfully represent certain facts of health-care interactions, yet not represent others. An insurance claim has some validity simply by nature of its existence—it is valid as a request for reimbursement for a health-care service. For example, a claim for a doctor visit for a covered patient will reliably be recorded, barring fraud or a data entry error. However, what will be less clear is what happened during the doctor visit. Attached to the doctor visit claim might be procedure codes and diagnosis codes that correspond to what services were rendered during the visit and what diagnoses were prompting the service. The patterns of health-care utilization over time and the associated procedure and diagnosis codes tell a story about a person, from which a great deal can be deduced about his or her health status, making the database valuable for research. However, the individual claims have limitations that translate into limitations of the database as a whole.

Weaknesses

Of course, insurance databases have important limitations as research tools, some of which can be reduced by enriching the data with added information from the provider or patient. The source administrative database was not designed for medical research, but rather to track financial transactions between patients' providers and the health insurance company. While the database may be adequate as a record of claims for reimbursement of services and payments made, important clinical details may be lacking. Examples of variables that claims data are typically lacking include disease severity markers, smoking, obesity, and physical activity.

A limitation of claims data with respect to characterizing exposure to drugs is a lack of information on patient adherence with the therapeutic regimen (see Chapter 42). A number of fields related to filling a prescription are provided in the pharmacy claim, such as dates filled, amount dispensed, and days' supply that allow for proxy measures of adherence and persistence.

Often health insurance databases are large, enabling the formation and follow-up of large cohorts.

These large cohorts can provide statistical power that makes even small effect sizes become apparent, increasing the difficulty of separating causal effects from spurious ones.

Commercial health insurance claims databases may be restricted to employed people and their dependents. The relative under-representation of people without employment can create methodologic difficulties for studies with hypotheses that are best studied in populations of elderly or unemployed. In the US, the elderly are under-represented in commercial insurance databases, with an extensive shift to Medicare (see Chapter 14) at age 65. Since drug effects in the elderly are often of great interest for pharmacoepidemiology, this aspect deserves careful assessment before proceeding with a research program in a commercial insurance environment. Commercial health insurance databases may not be directly generalizable to a population as a whole because of the under-representation of certain segments of the population. However, some relatively straightforward analytic methods (such as stratification) can address much of this limitation. Concerns regarding heterogeneity of treatment effects may persist, particularly if the hypothesized high-risk subgroup is not included or identifiable in the database, such as if unemployed persons are thought to be more vulnerable to an adverse effect being studied. Dimensions over which potential heterogeneity of treatment effect might be evaluated should be assessed within a commercial health insurance claims data environment that builds upon the claims with added clinical detail. Some relevant variables (e.g., clinical variables like left ventricular ejection fraction) might not be recorded within the electronically available data, and even where recorded, relevant levels of the variable (e.g., types of schizophrenia associated with the greatest disability) might not be present in the data.

The population covered by a commercial health insurance database may differ substantially from the geographic distribution of the population about which inferences are to be drawn. Geographic heterogeneity in the use of treatments, the occurrence of outcomes, and the association of the two may mean that regional composition of the population is important for addressing the pharmacoepidemiologic research questions. Accordingly, the composition of the database and a careful assessment of its inclusion of a population within a target region may be necessary. Geography also should be considered as a potential confounding variable. Related to geographical variation is variation arising from different health plans within a database. The differences in incentive structures (insurance benefit design) across the different health plans may lead to differences in insurance billing and coding across the plans with the potential to affect pharmacoepidemiologic assessments of exposures, outcomes or covariates. As such, the health plans within a broader health insurer could function as a confounder, effect modifier, or instrumental variable. Recognition of this potential source of heterogeneity and consideration of its effect on the results of a pharmacoepidemiologic study may be warranted.

Although the potential for unmeasured confounding to explain a research finding should be assessed when interpreting any observational research result (see Chapters 3 and 47), research conducted within a commercial health insurance claims data environment has particular susceptibility. Since such a study's source data are not designed for research, patient characteristics that are both associated with exposure and outcome might readily be hypothesized and might not be directly available within the database. Research designs that may be less susceptible to unmeasured confounding (self-controlled designs, active user designs, instrumental variable analyses; see Chapter 47) may be useful approaches, and sensitivity analyses may further serve to illuminate the degree of unmeasured confounding that would be needed to produce the observed result.

Free-living, employed people in the US tend to be mobile, and US commercial health insurance claims data environments reflect this mobility. The commercial health insurance claims database easily accommodates manifestations of this mobility related to patients seeking health care even far from home (such as on vacation) or from a range of different providers. However, the mobility of

employed people also leads to changes in health insurance either through a change in employment or the selection of a different employer-offered health insurer. Since a change in health insurer translates into entry or exit from the health insurer database, the dwell time of enrollees is affected by it. As a consequence, with the 25–30% annual turnover in enrollees, the average dwell time in a commercial health insurance data environment is typically 2 years or less. Although the average enrollment in the database may be less than 2 years, a specific cohort selected for a study may have an average enrollment duration of considerably more than 2 years. Further, even if the average duration of enrollment for a study cohort is 2 years, there will be a subset of the cohort that has longer enrollment, allowing for the assessment of longer-term effects in that subset (with assumptions regarding non-informative loss to follow-up). This average enrollment constraint should be considered when evaluating a health insurance database, and long-term exposure effects will generally be better evaluated elsewhere.

Commercial health insurance claims data do not include a patient history as an explicit file. A patient's medical history at any point in time will need to be inferred from the health-care interactions recorded in the insurance database. The accuracy of this medical history will depend on the time frame over which it is evaluated and the likelihood that a relevant medical condition will lead to a billable medical service within it. Often, this health history is derived from a baseline (before cohort entry) requirement that patients be continuously enrolled for some fixed time period over which health insurance claims are evaluated. Since completeness of this baseline health history depends on the time enrolled within a health insurer, care should be taken when the baseline period is not the same for all people included in a study. Further, baseline time periods should be distinguished from follow-up time, and only baseline time should contribute to covariates; time after the start of follow-up should not be included in the baseline period. Requiring a period of continuous baseline enrollment is generally necessary for developing a health history, but applying continuous enrollment that occurs after the start of follow-up can lead to the inclusion of immortal person-time (see Chapter 47) that might produce bias.[5] Analytic methods that update covariates in response to health events that occur during follow-up are available (e.g., marginal structural models) and might be applied if considered necessary for the research question.

Within a health insurance claims database, there will be some period of time that passes between the time a health service actually occurs and the time the claim for the service appears in the database and is available for research. This claims data lag has several components: (i) the time from when a provider performs the service until the claim for reimbursement is submitted; (ii) the time from when the claim is submitted until the claim is paid; and (iii) the time from when the claims is paid to when it is incorporated into a database accessible for research. The components of this lag vary considerably according to the category of claim, with pharmacy claims typically having the shortest lags (1–1.5 months), and hospital claims having the longest lags (3–6 months). This claims lag has the greatest implications for research on the most current data in a database. For data that is historical by a year or more, claims lag is unlikely to affect research.

Additional coding-related limitations common to all health insurance claims databases serve to limit internal validity, including diagnostic coding errors such as misspecification, miscoding, incorrect sequencing decisions, and clerical mistakes. Misspecification refers to errors associated with incomplete information from the provider, including providing the wrong diagnosis or failing to provide a diagnosis in the medical record. Miscoding refers to applying the wrong coding rules to the final diagnosis and is typically due to failure of the medical coder. Incorrect sequencing describes using diagnostic codes that are more applicable to secondary diagnoses as the primary diagnosis. Although the nature of these errors makes them rare within an insurance database overall (perhaps less than 0.1% of claims), it is possible that study selection criteria could increase their prevalence.[6]

Additional coding-related limitations can occur through other mechanisms, such as limits on the number of diagnosis fields allowed on a claim. For example, if a patient has five diseases, but the claim only allows the recording of three diagnoses, then information on a patient's clinical status may be incomplete. Further, certain health-care-related services may not be covered by a plan such as inexpensive drugs (might be less than the copayment amount) or not reimbursed by the health insurer (e.g., smoking cessation medications).

Particular applications

This section expands on two particular commercial insurance databases with which the authors are most familiar. Experience with other such databases suggests that the examples provided by these two should be generalizable to other commercial insurance databases. Other commercial health insurers, such as other Blue Cross/ Blue Shield plans, Humana, Aetna, CIGNA, and others, may have similar data and potential access to primary medical records.

The United Health data environment

UnitedHealth Group (www.unitedhealthgroup.com) is a national health-care company serving consumers, managers, and health-care professionals. Founded in 1974, the company serves more than 60 million persons through a continuum of health-care and specialty services. These services include point of service arrangements, preferred provider organizations, managed indemnity programs, Medicaid and Medicare managed care programs, senior and retiree insurance programs, and numerous other services.[7]

UnitedHealth Group-affiliated health plans presently reach across the United States and include both urban and rural representation. The members of the plans are predominantly employer-based groups but also include individuals from the Medicaid and Medicare populations. To serve these customers, the company arranges access to care with more than 400 000 physicians and 3300 hospitals. UnitedHealth Group-affiliated health plans are typically independent practice association models in which the providers are not employees of UnitedHealth, but a wide range of providers that that have agreed to accept patients who have UnitedHealth insurance. Health-care services are provided to patients and the providers submit claims for reimbursement to UnitedHealth.

UnitedHealth Group provides a link to affiliated health-care plans and their electronic administrative claims data. Typically, each affiliated health-plan contracts with a large network of physicians and hospitals to provide health-care services. These arrangements result in access to medical management information data reflecting a broad cross-section of the population with commercial health insurance, which provides UnitedHealth Group researchers and their collaborators with research opportunities.

Ingenix Research Database

The Ingenix Research Database, also termed the Normative Health Information (NHI) database, contains paid claims data for over 60 million fully-insured lives—people who have been enrolled for at least 1 day between 1994 and 2010. This longitudinal count of number of people reflects a cross-sectional size that varies considerably over these years and is currently approximately 12 million people.

Turnover of patients within the NHI database is consistent with expected turnover for commercially insured people. At approximately 30–35% per year, the effect on retention within a closed cohort can be estimated. Below are the numbers of fully-insured lives with 1, 2, 3, and 4 years of continuous medical and pharmacy eligibility from approximately 60 million:

- 1 year: 21 million
- 2 years: 23 million
- 3 years: 6 million
- 4+ years: 10 million.

Some of the data (approximately 50%) within the NHI are derived from health-care claims processing, also termed "administrative services only (ASO) members." For these health plans, UnitedHealth Group is not the insurer, and the data

are only used for research in de-identified form, so that only the electronic claims data are available for research.

Populations of children (almost 1 million in 2000), pregnant women (approximately 100 000 deliveries in 2009), and the elderly (over 200 000 were 65 years of age or older in 2000, and over 4 million in 2010) in the research databases are sufficient for many analyses. For specific applications, feasibility counts can be conducted to determine if numbers are adequate to address a particular research question.

Confidentiality and patient privacy

UnitedHealth Group adheres to strict confidentiality and patient privacy guidelines that meet the US Health Insurance Portability and Accountability Act (HIPAA) regulations (see Chapter 35). Research uses the minimum necessary protected health information, full documentation of such use is maintained, institutional review board/ privacy board approval with informed consent or waiver.

Some plans that have different financial incentives from the typical submission of a claim for reimbursement of health-care services (fee-for-service) may not have complete data. For example, if reimbursement to a specialist is capitated and there is no requirement to submit a bill for payment, that service may not be included as part of the database. This disadvantage may be addressed by excluding these few plans from data extraction for research studies. Another disadvantage is that certain variables are not available in the electronic claims databases, such as race/ ethnicity, cigarette smoking history, and obesity. If necessary for a specific study, this information may be ascertained through a review of medical records, to the degree that such information is present there.

The following examples provide empirical information on both the strengths and limitations of the data and present specific applications of the utilization of the UnitedHealth Group databases to address pharmacoepidemiologic research questions.

Examples

The value of the large size of the Ingenix Research Database becomes immediately apparent when studying a rare outcome such as Achilles tendon rupture,[8] or rhabdomyolysis,[9] or lymphoma.[10] Even these rare outcomes will occur numerous times within the database since the source population is so large. Even for a rare occurrence, the large number of cases that can be compiled allows for a rich description of the epidemiology to be established, including drugs as risk factors.

Examples of applications of this data source can be found in studies assessing whether findings from clinical trials are reproduced in observational settings,[11] or more classical pharmacoepidemiology, such as the effect of atomoxetine on stroke,[12] or seizure,[13] or the effect of the oral contraceptives on venous thromboembolism.[14]

Other examples include descriptions of patients who receive a particular drug in order to characterize the prescribing of that drug. This description can compare actual recipients of the drug to those expected based on the drug's indications to identify potential off-label prescribing, or this description might identify selective prescribing; variables with higher or lower prevalence among people who receive the drug that might need to be accounted for in order to mitigate confounding when making comparisons. Further, the patient characteristics of those prescribed the drug can be tracked longitudinally in order to assess the effect of interventions such as medication guides, Dear Doctor letters, and others, such as changes in cisapride dispensing following Dear Doctor letters[15] and the effect of cisapride use on outcomes (occurrence of serious ventricular arrhythmia with cisapride exposure).[16]

Classical pharmacoepidemiologic applications of the health insurance data might include research that associates drug exposure (ascertained through pharmacy dispensing records) with the occurrence of adverse outcomes (ascertained through medical claims and possibly validated through medical records), and comparative effectiveness research (see Chapter 32) can be conducted within commercial data environments in much the same way by associating exposure to a drug (or other intervention) with an effectiveness outcome. For example, the effectiveness of a rotavirus vaccine on the occurrence of gastroenteritis was evaluated.[17]

Additionally, effects of educational efforts or adherence to recommended monitoring guidelines can be evaluated.[18] Evaluations of educational interventions can extend beyond the prescribing process to include association with outcomes. For example, a study of the oral contraceptive ethinyl estradiol/ drospirenone included both adherence to recommended monitoring and the occurrence of hyperkalemic adverse events the monitoring was supposed to prevent.[19, 20]

It is possible to identify patterns of insurance claims that are associated with outcomes and to develop claims algorithms with measured accuracy. For example, hypersensitivity reactions following abacavir exposure can be identified using claims.[21] Sampled data collection can be targeted to address hypothesized confounding. For example, the potential for obesity or smoking to confound the association between the oral contraceptive ethinyl estradiol/ drospirenone use and venous thromboembolism was assessed in this way.[22] Finally, the effects of drug exposure during pregnancy can be assessed. For example, a study of the association of paroxetine and congenital malformations was conducted in this data source.[23]

The HealthCore data environment

HealthCore overview

HealthCore Inc. is an independently operating, wholly owned, research-focused subsidiary of WellPoint, Inc. WellPoint is one of the largest health benefits companies in terms of total enrollment in the United States, with medical enrollment in its affiliated health plans totaling approximately 33.5 million members as of September 30, 2010. WellPoint is an independent licensee of the Blue Cross and Blue Shield Association and serves its members as the Blue Cross licensee for California; the Blue Cross and Blue Shield licensee for Colorado, Connecticut, Georgia, Indiana, Kentucky, Maine, Missouri (excluding 30 counties in the Kansas City area), Nevada, New Hampshire, New York (as the Blue Cross Blue Shield licensee in ten New York City metropolitan and surrounding counties and as the Blue Cross or Blue Cross Blue Shield licensee in selected upstate counties only), Ohio, Virginia (excluding the Northern Virginia suburbs of Washington, DC), and Wisconsin.

The HealthCore data environment includes the HealthCore Integrated Research Database (HIRD^{SM}), the Integrated Research Network (IRN®) and other data collected from various sources. The HIRD^{SM} is a longitudinally integrated medical and pharmacy claims, enrollment, and electronic outpatient laboratory results database from health-care encounters of members of WellPoint's 14 affiliated health plans and one non-affiliated plan. Information regarding current and historical medical and pharmaceutical policies, benefit designs, formulary processes, and care management programs is accessed by HealthCore researchers within this research environment. The IRN® is a collaborative research community that engages providers in collaborative research with HealthCore.

The HealthCore Integrated Research Database (HIRD^{SM})

The HIRD^{SM} is a single health insurance database that contains approved claims integrated across data sources and data types (professional claims, facility claims, outpatient pharmacy claims, and enrollment information) and across years (from 2001 through the most recent calendar quarter). The specific US geographic regions represented in the HIRD^{SM} include the Northeast, Mid-Atlantic, Southeast, Midwest, Central, and the West.

Today, the HIRD^{SM} contains data for approximately 60.3 million lives with medical coverage and 43.8 million lives with both medical and pharmacy coverage at any point over the coverage period of the plans contained in the HIRD^{SM}. The discrepancy between the total current WellPoint affiliated health-plan membership and the number of lives represented in the HIRD^{SM} can be explained by several factors. First, the coverage period of the HIRD^{SM} is January 1, 2001 through September 30, 2010; 10 of the 14 plans' data includes the period from January 1, 2004 through September 30, 2010. Thus individuals who were members of a health plan contained in the HIRD^{SM} and whose data are available for research, but were no longer members as of September 2010, would still be represented in the HIRD^{SM}. Second, some health-plan members

represent groups whose data cannot be used for research, and therefore are not included in the counts. Examples of members whose data cannot be used for research at this time include members of state-sponsored programs such as Medicaid, members of the Federal Employee Program (FEP), and members of ASO customers who have not given permission to use their data for research. Formally, an ASO plan is a type of health plan where the employer or other group sponsor is financially responsible for paying plan expenses such as members' claims; the insurance company only provides administrative services. This is also called "self funded" or "self insured." Finally, the HIRD[SM] currently contains one health plan not affiliated with WellPoint. It is anticipated that the HIRD[SM] will continue to be expanded by the inclusion of additional non-WellPoint affiliated health plans.

Below are the numbers of researchable lives in the HIRD[SM] as of December 2010 with 1, 2, 3, and 4 years of continuous medical and pharmacy eligibility from the 43.8 million lives with both medical and pharmacy coverage:

- 1 year: approximately 28.5 million
- 2 years: approximately 19.1 million
- 3 years: approximately 12.2 million
- 4+ years: approximately 8.4 million.

Not all individuals have both medical and pharmacy coverage from the health plan. Research that examines the use of, or implications of prescription drugs, should use data from members who have both medical and pharmacy eligibility even though ensuring this requirement may have a moderate effect on reducing the study sample size. The rationale is that if the individual has both types of coverage, encounters in either setting would generally be identified in the claims data, which would be important for identification and measurement of the exposures of interest and the outcomes of interest. However, research examining the use or implications of use of vaccines, devices, or biologic products may not require the additional restriction of individuals with both medical and pharmacy coverage. Such an instance may be encountered when the particular product under investigation is only covered under the medical benefit (and therefore would not be identified in the pharmacy claims data). The advantage is that not requiring the additional constraint of pharmacy coverage will maximize the study sample size and statistical power. Caution should be used when deciding on eligibility requirements in the selection of the cohort to study, as some vaccines and biologic products may be administered in physician offices or dispensed in an outpatient pharmacy and thus requiring both types of coverage would be essential in order to maximize the capture of exposures.

Health plans included in the HIRD[SM] include lines of business such as health maintenance organizations (HMOs), point of service (POS), preferred provider organizations (PPOs), Consumer Directed Health Plans (CDHP), and indemnity plans. A PPO is a plan that allows members to choose any provider, but offers higher levels of coverage if members receive services from health-care providers in the plan's PPO network. These in-network providers have contracted with the health plan to provide services at prenegotiated reimbursement rates. An HMO is a type of health-benefits plan for which members are required to receive health care only from providers that are part of the HMO network. A primary care physician coordinates each member's health care. Services (except emergency care) performed by out-of-network providers aren't covered except under specified circumstances. A POS plan utilizes some of the features of PPO and HMO style plans; however, members do not make a choice about which system to use until the point at which the service is being used. These plans typically include levels of progressively higher patient financial participation as the patient moves away from the more managed features of the plan. A CDHP is a type of health plan designed to give members more control over their health care and how to spend their health-care dollars. These plans often include a high deductible and include a spending account such as a Health Reimbursement Arrangement (HRA), Health Savings Account (HSA), or Health Incentive Account (HIA). Indemnity plans are traditional insurance plans that reimburse for health-care services provided to members based on providers' bills submitted after the services are rendered.

Importantly, many variations of the above mentioned health plan types are often offered causing the distinction between the types to fade. It is quite challenging to adequately control for subtle differences between plans in the data, which may be directly impacting the utilization of services that are measured in the data.

Patient enrollment data, medical care, prescription drug use, and health-care utilization may be tracked for each patient throughout the course of their enrollment in the database. Diagnoses and procedures are identified by ICD-9 diagnostic, ICD-9 Procedure, Current Procedural Terminology (CPT), and Health Care Financing Administration Common Procedure Coding System (HCPCS) codes for both outpatient and inpatient visits/ stays. Outpatient pharmacy claims are captured by National Drug Codes (NDCs), which can then be translated to broader classification systems such as Generic Product Identifier (GPI) codes. This type of classification systems can be more meaningful than NDCs in the sense that drug classes and subclasses can be defined and grouped regardless of manufacturer. Physician, specialist, and emergency room visits, as well as hospital stays, are captured in the database through ICD-9 diagnostic, CPT procedures codes, HCPCS, Uniform Billing Code of 1992 (UB-92) revenue codes (e.g., room and board), and place of service codes. Information on physician specialty is also retained in the database at a level consistent with National Committee for Quality Assurance (NCQA) credentialing requirements.

Demographic characteristics available in the HIRD^{SM} include date of birth and gender from enrollment data. In addition, each member represented in the HIRD^{SM} is associated with a unique encrypted identification (ID) number. Standard procedures entail retaining each member's unique ID number throughout the period of plan eligibility. Race and ethnicity are not available in the administrative claims database; however these variables, in addition to socioeconomic status, can be obtained through several direct and indirect mechanisms. Direct mechanisms include abstraction of medical charts or via patient survey. Indirect methods include imputation based on US Census data and member zip code and/or US Census tract information. Furthermore, family clusters can be identified and linked via the combination of subscriber and dependent ID codes.

Pregnancies can also be identified in the HIRD^{SM} using ICD-9 diagnosis and CPT procedure codes.[24] While identifying pregnancy in claims is fairly uncomplicated, determining the beginning of gestation, in order to identify prenatal exposures, is more challenging and may involve identifying deliveries using procedure codes and then "counting back" the weeks assuming a normal gestation. This method, however, would miss early deliveries, spontaneous abortions, or other pregnancy termination. Medical chart review can be used to determine accuracy of algorithms used to identify pregnancies. Additionally, claims for newborns may initially be paid through the mother's ID until the newborn has an ID assigned. Linkage between mother and child improves the ability to allocate the child's claims to the child; however, some claims that are identified during this period may not be clearly assigned to mother or child. For example, a neonatal intensive care unit stay or a circumcision procedure clearly belongs to the child while an antibiotic dispensed could belong to either mother or child.

The HIRD^{SM} also contains diagnostic laboratory testing results from certain reference laboratories for individuals receiving outpatient laboratory services. The accessible laboratory data covers between 20 and 30% of outpatient diagnostic testing, depending upon the region. Data include full ranges of hematologic, chemistry, immunologic, and microbiologic (including antibiotic sensitivity results) testing. As of December 2010, outpatient laboratory results were contained in the HIRD^{SM} for approximately 10.3 million lives with medical coverage and 8.0 million lives with medical and pharmacy coverage. It is possible that individuals who have laboratory results available in the HIRD^{SM} may also receive laboratory testing at other outpatient facilities that do not provide results to HealthCore or in the inpatient setting. These results would not be included in the HIRD^{SM}; however, evidence that tests occurred can be measured with CPT codes regardless of laboratory provider.

The HIRD[SM] is updated on a monthly basis with as short as a 1 to 2-month lag. However, medical claims traditionally used by HealthCore for research may have up to a 4-month lag in order to use the most accurate data available accounting for claim reversals, time in submitting claims by providers, time in paying claims by plans, and time for HealthCore personnel to extract, transform, and load data into the HIRD[SM]. The lag time for pharmacy data is shorter since typical claim submission and adjudication occurs instantaneously with dispensing.

HIRD[SM] enrichment

As a HIPAA Privacy Business Associate of the health plans whose data it receives, HealthCore is able to populate the HIRD[SM] with fully identifiable data, which allows for the ability, with appropriate HIPPA Waivers of Authorization and/or institutional review board approval, to abstract inpatient and outpatient medical records, to identify and contact providers and members for survey research, and to link data to national vital records and registries.

HealthCore abstracts medical/hospital charts using identifiers obtained from HealthCore's administrative database. Clinical information abstracted from the medical/hospital charts is entered into a study database maintained by HealthCore with a masked identifier so that it can be matched with corresponding claims data without the use of individually identifiable information, such as patient name or medical record number.

In addition to medical chart review, individuals whose data are included in the HIRD[SM] have been contacted and invited to participate in survey research, on a study by study basis. Linking the survey responses to administrative claims data allows for determining baseline demographic characteristics, co-morbidities, mild adverse effects (such as from prescription drug or biologic products) not reported to a health-care provider, prescription medication utilization, and differences in health-care costs and utilization between cohorts.

Data from the HIRD[SM] have also been linked to information on mortality. Two sources of mortality data that have been used to integrate such information with the HIRD[SM] for specific projects have been the National Death Index (NDI) and Social Security Administration's (SSA) Death Master File (DMF). Direct identifiers such as patient name and date of birth are commonly used to conduct matches between the two data sources. HealthCore has a subscription to the DMF, which includes more than 65 million individual recorded deaths reported to the SSA with the following information (if data are available to the SSA): social security number, name, date of birth, date of death, state or country of residence, zip code of last residence, and zip code of the lump sum payment. The most "up-to-date" information on death is available from the SSA. Cause of death cannot be determined using the DMF. Cause of death can be obtained from the NDI-plus data provided by the National Center for Health Statistics (NCHS). Disadvantages of the NDI include its long lag time for both data availability (about 2 years) and several month evaluation and approval period before NDI data will be released. In addition, the HIRD[SM] has been linked to registries, such as state cancer registries and immunization registries.

The Integrated Research Network (IRN®)

The IRN is the second component of HealthCore's data environment. The IRN is a national collaborative research community of practicing physicians and health-care facilities that can serve as partners in HealthCore's research activities.

The IRN enables efficient research as well as providing access to a very large provider network to enable enrichment of HIRD[SM] claims data with provider data from the medical record (paper-based and electronic), supplemental clinical data prospectively collected at the provider site, or valuable clinical insight and specialty expertise from the participating physicians that can serve to inform the research. In total, through the health-plan provider networks, HealthCore has access to more than 276 000 physicians, 1600 facilities, and 62 000 pharmacies. This network serves as the population from which HealthCore has identified provider groups that treat a large portion of individuals represented in the HIRD[SM] who have claims evidence of chronic conditions that have been studied by HealthCore. Currently, the IRN includes

approximately 3900 physicians who treated approximately 1.5 million individuals represented in the HIRD[SM].

The goals of the IRN are: (i) to perform real-world research to evaluate the effectiveness and safety of drugs, treatments, and health-care interventions and to provide meaningful medical evidence back to the IRN participants to help physicians in understanding treatment patterns and in providing appropriate treatment to their patients; (ii) to work with physicians and other health-care experts to infuse clinical relevance and applicability into health outcomes, comparative effectiveness, and epidemiologic research; and (iii) to electronically combine clinical information with the HIRD[SM] to enhance and/or enable prospective comparative effectiveness, safety, and health outcomes studies.

Healthcare Safety Sentinel System (HSSS)

The Food and Drug Administration Amendment Act of 2007 (FDAAA-PL 110-85) mandated the Food and Drug Administration (FDA) to establish a prospective drug safety surveillance system using electronic health-care data.[25] FDA's response to this mandate, the Sentinel System (see Chapter 30), is utilizing data arising from large, commercial claims data environments. It is in this context that HealthCore has developed the Healthcare Safety Sentinel System (HSSS). The HSSS is a targeted, active, prospective safety monitoring system that leverages HealthCore's data environment, including the HIRD[SM], and is intended to be interoperable with and complementary to FDA's Sentinel Initiative surveillance system. The HSSS was launched in 2009 with the objective of enabling rapid, early detection of drug safety concerns. The HSSS includes a tested, repeatable, end-to-end process, which includes the identification of an initial safety concern, assembly of a clinical evaluation research team, formulation of an investigative plan, semiautomated cohort identification, data extraction and propensity score matching, quantitative risk analysis and application of signal identification algorithms, report generation and evaluation, decision point for continued monitoring (or not), legal evaluation, and development of a communication plan. Currently, the HSSS uses a propensity score-match cohort design to rapidly monitor specific drug–adverse event concerns that are initiated by an external entity (e.g., drug manufacturer, FDA, health plans, etc.) that would engage with HealthCore to monitor the exposure–event combination. The ability to validate events with medical record review is a central feature of the HSSS.

As part of the continued development of the HSSS, between 2008 and 2010 HealthCore validated eight events of medical concern in order to: (i) determine if data algorithms already created have comparable positive predictive value (PPV) to that of other electronic databases; (ii) identify algorithms for medical concerns of interest identified by the FDA; and (iii) validate and create algorithms for medical events of concern that apply to large populations of patients or are part of ongoing safety work. Algorithms include combinations of diagnosis, procedure codes, and place of service and have been developed and validated for myocardial infarction (MI), ischemic stroke, severe upper gastrointestinal (UGI) bleed, thrombotic thrombocytopenic purpura (TTP), Stevens–Johnson syndrome (SJS), severe hypoglycemic events, and neutropenia.[26–31] These validations have been conducted using abstracted paper-based or electronic medical records linked to HIRD[SM] claims data. Typically, such validation studies using paper-based medical records include claims identification of possible adverse events of interest, development of a chart abstraction form designed to collect necessary clinical information for a physician specialist to determine case status, abstraction of data elements form the medical records by professional nurse abstractors, review and adjudication of all data from records by physicians with specialty in the condition of interest, and calculation of the PPV of the claims algorithm that was used to identify the case.

Confidentiality and patient privacy

HealthCore receives data from health plans and fully understands the responsibility of data privacy stewardship. Use and disclosure of data is in compliance with current regulations related to data privacy and security. These include the US HIPAA

Privacy and Security Rule, 45 CFR 160, and 164, "The Common Rule," 45 CFR 46, the FDA Protection of Human Subjects Regulations, 21 CFR Parts 50 and 56 and "The Part 2 Rules", 42 CFR 2 (see also Chapter 35). Additionally, state regulations, which further limit the use and disclosure of data, are reviewed at a project level for compliance. Requests for the use and disclosure of data are reviewed by HealthCore's Regulatory Compliance Office for adherence to the appropriate regulations.

Applications using the HealthCore data environment

HealthCore's research experience spans safety and epidemiology as well as health services research, including health economics and outcomes research (HEOR) and comparative effectiveness research (CER). The following describes a few examples of studies that have been conducted using data from HealthCore's data environment.

Drug and vaccine safety research

1 In response to a 2008 FDA requirement for new labeling on anticonvulsant medications that warned of the increased risk of suicidal thoughts and behavior, based on a meta-analysis of placebo-controlled trials which lacked conclusive evidence about safety at the individual agent level, a retrospective cohort study using the HIRD^{SM} evaluated such risks associated with individual anticonvulsants.[32] Data from the HIRD^{SM} were linked with death data from the SSA DMF and NDI-plus to determine date and cause of death for individuals 15 years of age or older with pharmacy claims for anticonvulsant medications between July 2001 and December 2006. A total of 297 620 new episodes of anticonvulsant treatment were included in the analyses. The results suggested that increased risk of suicidal acts or violent death was associated with use of gabapentin, lamotrigine, oxcarbazepine, and tiagabine compared with topiramate.

2 Four health plans from the HIRD^{SM} were used to evaluate the risk of acute myocardial infarction (AMI), acute heart failure (AHF), or all-cause death among pioglitazone and rosiglitazone-treated patients.[33] Using claims data between 2001 and 2005 and the NDI-plus for 36 628 new users of pioglitazone or rosiglitazone, and a propensity score-matched retrospective cohort design, this study observed no significant difference between matched groups for risk of the composite event of AMI/AHF/death.

3 The HIRD^{SM} is also being used in an observational study to examine the cardiovascular safety of the following commonly-used medications to treat attention deficit hyperactivity disorder (ADHD): amphetamines, methylphenidate, and atomoxetine. Primary outcomes of interest are: (i) sudden death/ ventricular arrhythmia, (ii) stroke, (iii) myocardial infarction, and (iv) stroke or myocardial infarction as a composite outcome. Secondary outcomes are: (i) all-cause death, (ii) non-suicide death, and (iii) non-accident death.[34]

4 Vaccines have also been studied using the HIRD^{SM}. As part of a multicenter post-marketing safety study using a distributed network approach that included the HIRD^{SM}, patients with exposure to the meningococcal vaccine were identified in claims data and followed longitudinally for claims evidence of Guillain–Barré syndrome (GBS). Full-text medical records were abstracted and reviewed to adjudicate the case status member with evidence of GBS (ICD-9 357.0).[35] HealthCore participated as one of five research organizations closely affiliated with US health insurers. Study results are forthcoming.

Health services research, HEOR, comparative effectiveness

1 Short *et al*. linked HIRD^{SM} data from one health plan to patient medical records in order to measure ethnic disparities in medical care among commercially insured women newly diagnosed with breast cancer.[36] Individuals newly diagnosed breast cancer were identified by breast cancer ICD-9 codes observed in medical claims. Member and physician office zip codes that matched regions with more than 50% minorities according to 2000 census tract were used. Results suggested that African-American women were diagnosed at later stages than White women and higher percentages of

white women received antiestrogen therapy than African-American women, after adjustment for important confounding variables using multivariate regression. This study revealed the presences of ethnic differences in management of breast cancer and suggests that development of policies aimed at reducing these disparities would have the potential to improve equity of care.

2 Another study, which HealthCore recently completed, led to a policy change regarding treatment for asthma within the health plans.[37] The study evaluated the impact of asthma medication on patient-reported asthma control and used claims data integrated with patient surveys. Compared to inhaled corticosteroids (ICs), leukotriene modifiers (LMs) were associated with lower odds of inpatient stays and emergency department (ED) visits, lower odds of using six or more short-acting beta-agonist containers, and higher annual costs. However, among the relatively small subgroup of members adherent on their medications, LMs were associated with higher odds of inpatient and ED visits, lower odds of using six or more short-acting beta agonist containers, and higher annual costs. Both had a comparable impact on patient-reported quality of life. Results suggested that when patients cannot achieve medication adherence on their asthma medications, LMs may offer a reasonable alternative even though they are higher cost and typically in a higher tier structure than ICs. These results have aided health plans in formulary decision making in order to help control asthma better.

3 In addition to evaluations of the effectiveness and economics associated with treatment strategies, the HIRD^{SM} has also been used to evaluate novel health information technologies. Daniel *et. al*. evaluated the use of a payer-based electronic health record (P-EHR), which is a clinical summary of a patient's medical and pharmacy claims history, in an ED on length of stay (LOS) and plan payments.[38] The P-EHR was implemented by a large, urban ED in partnership with a HIRD^{SM} health plan for widespread use for health-plan members presenting to the ED. The evaluation used a retrospective study design with a sample of historical control ED visits. Claims data from the HIRD^{SM} linked to supplemental electronic hospital data, including ED LOS measured in minutes, the emergency severity index (ESI) measure used at triage, the census of the ED at the time of each ED visit, and whether or not the visit resulted in hospital admission or discharge to home. Analyses evaluated the impact of using the P-EHR in the ED setting on study outcomes using multivariate regressions and the non-parametric bootstrap. After covariate adjustment, among visits resulting in discharge (ED-only), P-EHR visits were 19 minutes shorter (95% confidence interval [CI] = 5–33 minutes) than non-P-EHR visits. Among visits resulting in hospitalization, the P-EHR was associated with a 77-minute reduction in LOS (95% CI: 28–126) compared to non-P-EHR visits. The P-EHR was associated with $1560 (95% CI: $43–$2910) savings in total plan expenditures for hospitalized visits. This study suggested that health information technology that provides ED physicians with medical and prescription drug histories, based on health-plan claims data, may lead to improvement in the efficiency of care among all visits and lower costs of care for visits that result in hospitalization.

The future

Postmarketing surveillance and pharmacoepidemiologic research studies are, by necessity, conducted in the context of a rapidly changing health-care environment. Looking to the future, this dynamic environment has implications for the nature of the data obtained, the ability to obtain information, and the characteristics of the general and specific populations that form the basis of these types of studies. Changes in the health-care system, given the concern with the rising cost of health care, will continue to impact the health-benefits structure, including pharmacy-benefits coverage. As the pharmacy-benefit structure is broadened to include coverage across a larger number of defined tiers (e.g., brand, generic, and not on the preferred drug list), certain transactions may not be captured in the pharmacy files if the cost is lower than the copayment. For example, some retail pharmacies offer select, generically available medications for a

few dollars ($4) and these dispensings may not be captured by the health insurer.

Claims database owners will seek to expand (i.e., add lives through inclusion of more health plans) and these will become incorporated into the research database, so that these research environments will continue to grow in size. Further, automation and generation of electronic tools to enable efficient data linkages, feasibility assessments, and, in some cases, execution of safety monitoring protocols will commence. It will be critically important that such automation does not occur without the deliberate and continued validation of such processes, given the limitations of claims data. As such, these databases will continue to have enhancements made to them that improve their ease of use, speed of access, reliability, and efficiency, all while shortening the lag between when a service actually occurs and when the claim for it is incorporated into the database for research. As pharmacoepidemiologic research continues to use commercial health insurance claims data, more will be understood about how the data reflect what happened to the patient throughout the course of health-care utilization.

Linkages between commercial health insurance claims and new electronic data sources from the health-plan provider network, including electronic medical records, detailed inpatient data, and personal health records containing such information as member health risk assessments, will continue and be available for future research.

On October 1, 2013, health-care organizations in the US will begin mandatory migration from using ICD-9 to ICD-10 for medical diagnosis and inpatient procedure codes (www.cms.gov/ICD10/). This change will increase the number of codes from 17 000 to more than 141 000, including increasing the number of outpatient diagnosis codes from 13 500 to 68 000. There are many potential benefits, including better description and categorization of clinical conditions and services rendered by providers within claims databases. These changes will, however, lead to significant changes within claims databases and will require a great deal of scrutiny and revalidation prior to use in research.

The importance of pharmacoepidemiology and the ability to conduct postmarketing studies will become increasingly critical in the future due to a number of factors relating to changing characteristics of the source populations. First, changes attributable to the aging of populations will augment the need for a better understanding of the use of prescribed medications in the older population, including variations in metabolism and appropriate dosages. Second, as the population ages, more populations are likely to receive a larger number of medications due to higher prevalences of chronic diseases, raising polypharmacy questions. Increased use of over-the-counter drugs, alternative medications, and devices also will need to be explored. Third, developments in the biotechnology field are resulting in new and expensive, but potentially more effective products, whose adverse effects need to be addressed. Finally, the commercial incorporation of pharmacogenomic data to target appropriate patients should increase both the effectiveness and safety profiles of new therapies (see Chapter 34).

Lastly, as the demand for real-world evidence grows, researchers are increasingly looking to electronic health-care databases as sources for this work. Initiatives designed to stimulate development of these resources are becoming widespread, including FDAAA (PL 110-85),[25] the FDA's Sentinel Initiative (see Chapter 30), and the magnitude of funding for comparative effectiveness research under the American Recovery and Reinvestment Act of 2009 (ARRA), with additional funding now forthcoming from the new Patient-Centered Outcomes Research Institute (see Chapter 32). Such initiatives have the potential for accelerating progress in the repurposing of these data for research. There are, however, risks associated with many of these well-intended efforts. In the quest for more available, affordable sources of data, it cannot be forgotten that much of the testing and validation of these sources has not yet been completed. The FDA Mini-Sentinel Initiative (see Chapters 12 and 30) was developed with this risk in mind. Mini-Sentinel uses a collaborative federated model where each participant contributing data has established an effective research data envi-

ronment and is an active collaborator in the initiative. However, some of the initiatives aiming to aggregate data from multiple sources, such as the All Payer Claims Databases, are effectively separating the data from its original source system. With this separation comes the increased risk of misinterpretation of the data and the risk of errant attribution of cause and effect. With this in mind, all research deriving from a health insurance claims data source should be interpreted with attention to the potential limitations outlined in this chapter.

References

1. Shatin D, Drinkard C, Stergachis A. UnitedHealth Group. In: Strom BL, ed. *Pharmacoepidemiology*, 3rd edn. Chichester: John Wiley & Sons, 2000, pp. 295–305.
2. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epi* 2005; **58**: 323–37.
3. Shatin D. Organizational context and taxonomy of health care databases. *Pharmacoepidemiol Drug Saf* 2001; **10**: 367–71.
4. Selby JV. Linking automated databases for research in managed care settings. *Ann Intern Med* 1997; **127**: 719–24.
5. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol* 2008; **167**: 492–9.
6. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997; **127**: 666–74.
7. Tuckson R. Costs of health care administration in the United States and Canada. *N Engl J Med* 2003; **349**: 2462–3.
8. Seeger JD, West WA, Fife D, Noel GJ, Johnson LN, Walker AM. Achilles tendon rupture and its association with fluoroquinolone antibiotics and other potential risk factors in a managed care population. *Pharmacoepidemiol Drug Saf* 2006; **15**: 784–92.
9. McAfee AT, Ming EE, Seeger JD, Quinn SG, Ng EW, Danielson JD, *et al*. The comparative safety of rosuvastatin: a retrospective matched cohort study in over 48,000 initiators of statin therapy. *Pharmacoepidemiol Drug Saf* 2006; **15**: 444–53.
10. Schneeweiss S, Doherty M, Zhu S, Funch D, Schlienger RG, Fernandez-Vidaurre C, *et al*. Topical treatments with pimecrolimus, tacrolimus and medium- to high-potency corticosteroids, and risk of lymphoma. *Dermatology* 2009; **219**: 7–21.
11. Loughlin J, Quinn S, Rivero E, Wong J, Huang J, Kralstein J, *et al*. Tegaserod and the Risk of Cardiovascular Ischemic Events: An Observational Cohort Study. *J Cardiovasc Pharmacol Ther* 2010: **15**: 151–7.
12. Holick CN, Turnbull BR, Jones ME, Chaudhry S, Bangs ME, Seeger JD. Atomoxetine and cerebrovascular outcomes in adults. *J Clinical Psychopharmacology*, 2009; **29**: 453–60.
13. McAfee AT, Holdridge KC, Johannes CB, Hornbuckle K, Walker AM. The effect of pharmacotherapy for attention-deficit/hyperactivity disorder on risk of seizure in pediatric patients. *Current Drug Safety* 2008; **3**: 123–31.
14. Seeger JD, Loughlin J, Eng PM, Clifford CR, Cutone J, Walker AM. Risk of thromboembolism in women taking ethinylestradiol/drospirenone and other oral contraceptives. *Obstet Gynecol* 2007; **110**: 587–93.
15. Weatherby LB, Nordstrom BL, Fife D, Walker AM. The impact of wording in "Dear Doctor" letters and in black box labels. *Clin Pharmacol Therap* 2002; **72**: 735–42.
16. Enger C, Cali C, Walker AM. Serious ventricular arrhythmias among users of cisapride and other QT-prolonging agents in the United States. *Pharmacoepidemiol Drug Saf* 2002; **11**: 477–86.
17. Wang FT, Mast TC, Glass RJ, Loughlin J, Seeger, JD. Effectiveness of the pentavalent rotavirus vaccine in preventing gastroenteritis in the United States. *Pediatrics* 2010; **125**: e208–e213.
18. Funch D, Norman H, Wong J, Bennett R, Hillier J, Al-Sabbagh A, *et al.* Adherence to Recommended Dosing and Monitoring for Mitoxantrone in Patients with Multiple Sclerosis: A Healthcare Claims Database Study Supplemented with Medical Records—The RETRO Study. *Pharmacoepidemiol Drug Saf* 2010; **9**: 448–56.
19. Eng PM, Seeger JD, Loughlin J, Oh K, Walker AM. 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**: 101–7.
20. Loughlin J, Seeger JD, Eng PM, Foegh M, Clifford, CR, Cutone J, *et al*. Risk of hyperkalemia in women taking ethinylestradiol/drospirenone and other oral contraceptives. *Contraception* 2008; **78**: 377–83.
21. Nordstrom BL, Norman HS, Dube TJ, Wilcox MA, Walker AM. Identification of abacavir hypersensitivity in health care claims data. *Pharmacoepidemiol Drug Saf* 2007; **16**: 289–96.

22. 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. *Pharmacoepidemiology and Drug Saf* 2008; **17**: 297–305.
23. Cole JA, Ephross SA, Cosmatos IS, Walker AM. Paroxetine in the first trimester and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1075–85.
24. Kuriya B, Hernández-Díaz S, Liu J, Bermas BL, Daniel G, Solomon DH. Patterns of drug use during pregnancy in rheumatoid arthritis. *Arthritis Care Res* 2011; **63**: 721–8.
25. US Food and Drug Administration, Department of Health and Human Services. Food and Drug Administration Amendments Act of 2007. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_public_laws&docid=f:publ085.110 (accessed June 2011).
26. Wahl PM, Rodgers K, Schneeweiss S, Gage BF, Butler J, Wilmer C, *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**: 596–603.
27. Wahl PM, Terrell DR, George JN, Rodgers JK, Uhl L, Cataland S, *et al*. Validation of claims-based diagnostic codes for idiopathic thrombotic thrombocytopenic purpura in a commercially-insured population. *Thromb Haemost* 2010; **103**: 1203–9.
28. Rosenman MB, Wahl PM, Daniel G, Overhage JM, McGuire P, Thompson D, *et al*. *Electronic Medical Records System Validation of Claims-Based Queries for Acute Kidney Injury*. Poster presented at the National Kidney Foundation (NKF) Spring Clinical Meetings, Nashville, TN, March, 2009.
29. Daniel GW, Eisenberg DF, Jones JK, Goehring EL, Winters P, Levin J, *et al*. Validation of a claims-based diagnostic code for Stevens–Johnson syndrome (SJS) in a commercially-insured population. Poster presented at the International Society for Pharmacoepidemiology (ISPE), *26th International Conference of Pharmacoepidemiology and Therapeutic Risk Management* (ICPE). Brighton, England, UK, August 19–22, 2010. http://www.pharmacoepi.org/meetings/26thconf/scientific_program_pt2.pdf (accessed June 2011).
30. Eisenberg DF, Daniel GW, Subramanian U, Breen T, Bohn RL, Wahl PM, Herman WH. Validation of claims-based diagnostic codes for severe hypoglycemic events (SHEs) in a commercially-insured population. Poster presented at the International Society for Pharmacoepidemiology (ISPE) *26th International Conference of Pharmacoepidemiology and Therapeutic Risk Management* (ICPE). Brighton, England, UK, August 19–22, 2010. http://www.pharmacoepi.org/meetings/26thconf/scientific_program_pt2.pdf (accessed June 2011).
31. Kim SY, Solomon DH, Lui J, Chang G, Daniel GW, Schneeweiss S. Accuracy of identifying outpatient neutropenia diagnoses in claims data. Poster presented at the International Society for Pharmacoepidemiology (ISPE) *26th International Conference of Pharmacoepidemiology and Therapeutic Risk Management* (ICPE). Brighton, England, UK, August 19–22, 2010. http://www.pharmacoepi.org/meetings/26thconf/scientific_program_pt2.pdf (accessed June 2011).
32. Patorno E, Bohn RL, Wahl PM, Avorn J, Patrick AR, Liu J, Schneeweiss S. Anticonvulsant medications and the risk of suicide, attempted suicide, or violent death. *JAMA* 2010; **303**: 1401–9.
33. Wertz DA, Chang CL, Sarawate CA, Willey VJ, Cziraky MJ, Bohn RL. Risk of cardiovascular events and all-cause mortality in patients treated with thiazolidinediones in a managed-care population. *Circ Cardiovasc Qual Outcomes* 2010; **3**: 538–45.
34. Hennessy S, Schelleman H, Daniel GW, Bilker WB, Kimmel SE, Guevara J, *et al*. Cardiovascular safety of attention deficit hyperactivity disorder medications: rationale for and design of an investigator-initiated observational study. *Pharmacoepidemiol Drug Saf* 2010; **19**: 934–41.
35. Velentgas P, Bohn RL, Brown JS, Chan A, Gladowski P, Holick CN, *et al*. A distributed research network model for post-marketing safety studies: the Meningococcal Vaccine Study. *Pharmacoepidemiol Drug Saf* 2008; **17**: 1226–34.
36. Short LJ, Fisher MD, Wahl PM, Kelly MB, Lawless GD, White S, *et al*. Disparities in medical care among commercially insured patients with newly diagnosed breast cancer: opportunities for intervention. *Cancer* 2010; **116**: 193–202.
37. Tan H, Sarawate C, Singer J, Elward K, Cohen RI, Smart BA, *et al*. Impact of asthma controller medications on clinical, economic, and patient-reported outcomes. *Mayo Clin Proc* 2009; **84**: 675–84.
38. Daniel GW, Ewen E, Willey VJ, Reese CL, Shirazi FM, Malone DC. Efficiency and economic benefits of a payer-based electronic health record in an emergency department. *Acad Emerg Med* 2010; **17**: 824–33.

CHAPTER 14
US Government Claims Databases

Sean Hennessy,[1] *Cristin Palumbo Freeman,*[1] *and Francesca Cunningham*[2]

[1] Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[2] Pharmacy Benefits Managements Services, Department of Veterans Affairs, Center for Medication Safety, Hines, IL, USA

Introduction

The United States (US) government funds healthcare services for certain segments of the population via a number of programs. Data from three of these programs have been used extensively for pharmacoepidemiologic research, and are the focus of this chapter. These three programs are Medicaid, Medicare, and the Department of Veterans Affairs (VA) Heath Care System. These programs and their resulting data differ substantially with regard to populations covered, benefits provided, and data available. Medicaid data have been used for pharmacoepidemiologic research since 1980, while VA data and especially Medicare data have been useful for pharmacoepidemiology more recently. This chapter describes these data sources, their strengths and limitations, and gives examples of how these data have been used in pharmacoepidemiologic research. We then discuss the future of these databases.

Description

Medicaid

Description of the Medicaid program

Medicaid was established in 1965, and is administered separately by each state or territory that offers Medicaid coverage, with federal oversight from the Centers for Medicare and Medicaid Services (CMS). Thus, there are 56 separate Medicaid programs, which have common features, but also differences, described below. These programs are funded jointly by the federal government and individual state governments. The federal share of funding varies from 50 to 76% depending on the individual state's average per capita income. Medicaid functions as a payer rather than a provider of health services. Each state establishes its own eligibility rules within general federal mandates. Most states also have "state-only" programs that pay for medical services for specific categories of people not qualifying under any federally specified category.[1] However, states cannot use federal funds to pay for state-only programs without a waiver from CMS.

Medicaid is currently the largest US government-funded health care program, providing coverage to 58 million US citizens and lawfully admitted immigrants, of whom approximately 75% are either low income pregnant women or members of low income families with children, with the remaining 25% consisting of chronically disabled or low-income elderly persons, including some who also receive Medicare benefits.[2] Medicaid does *not* currently provide coverage for even the poorest individuals *unless* they belong to one of these specifically designated groups. However, eligibility rules will be changing over the next few years following passage of the Patient Protection and Affordable Care Act (PPACA) in 2010. This law is expected to extend Medicaid coverage to an additional 16 million people by 2019, with low-income childless adults

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

under age 65 being the largest group to become newly eligible.[3]

The services covered by Medicaid vary by state, within federal mandates. Certain services are mandatory, including inpatient and outpatient hospital services, and physician services. Although outpatient prescription drug coverage is not mandatory, all Medicaid programs provide such coverage for at least some enrollment categories.[4] With certain exceptions (e.g., drugs for anorexia, weight gain, fertility, etc.), Medicaid programs are required to cover all drugs manufactured by companies that have entered into a federal rebate agreement, which includes all or nearly all manufacturers. However, programs can require prescribers to obtain prior authorization before the prescription will be covered. With the enactment of the Medicare drug benefit in 2006, outpatient prescription drugs for Medicaid–Medicare dual enrollees are now covered by Medicare rather than Medicaid. Most states pay the cost of drugs covered by the Medicaid program but not Medicare.[5] Medicaid also covers a large proportion of all nursing home care in the US, with over 40% of long-term care costs billed to the program.[6]

Most Medicaid programs include beneficiaries who receive services on a fee-for-service basis as well as those enrolled in capitated plans. In fee-for-service plans, health care providers bill Medicaid for specific goods and services provided, such as visits, hospitalizations, and prescription drugs. In capitated managed care plans, an insurance company is paid a certain amount per person per time period (e.g., month) to cover all or specific aspects of that enrollee's health care. Importantly for researchers, the degree of completeness of encounter information for patients in capitated plans is believed to vary among plans, although this has not been formally studied.

Health care providers participating in Medicaid must accept Medicaid payment as payment in full, although states may impose nominal deductibles, co-insurance, or co-payments on some Medicaid beneficiaries for certain services. For example, most state Medicaid programs allow pharmacies to charge patients a co-payment, which ranges from $0.50 to $5.00, for each outpatient prescription, with some states setting monthly limits for co-payments (e.g., Montana limits co-payments to $25 per month).[4] As noted, because of federal law, state Medicaid programs must cover all prescription drugs with a few exceptions of classes for which the coverage decision is left to the states (e.g., fertility treatments, benzodiazepines, barbiturates).

Characteristics of Medicaid recipients

In 2008 (the most recent year for which demographic data are available), 58.2 million persons, or 19% of the US population, received health care services covered by Medicaid. Baseline characteristics of the Medicaid population can be seen in Table 14.1. Children, females, and non-whites are overrepresented in Medicaid.

Sources of Medicaid data for research

CMS is the major source of Medicaid data for researchers. It receives data from individual state Medicaid programs, and performs extensive editing, range checks, and comparisons with previous data from that state when preparing research files, known as Medicaid Analytic Extract (MAX) files. Anomalies in the data are either reconciled with the state or published in an anomaly report, which is available to researchers. Crude data provided by states through the Medicaid Statistical Information System (MSIS) are also available, but do not undergo the same quality assurance checks as the MAX data. There is currently a lag of approximately 3 to 4 years between the end of a calendar year and when MAX data from that year become available.

The CMS' Research Data Assistance Center (ResDAC), operated through a contract with the University of Minnesota School of Public Health, provides free assistance to academic, government, and non-profit researchers in obtaining and using Medicaid and Medicare data. ResDAC maintains a website of information on Medicaid and Medicare data (http://www.resdac.umn.edu/), conducts workshops and seminars, and provides individual technical assistance to researchers, including obtaining prices for data from CMS, assisting in preparation

Table 14.1 Demographic characteristics of the Medicaid, Medicare, and Veterans Health Administration populations[56–59]

Characteristic		Medicaid (2008)‡		Medicare (2009)						Veterans Health Administration (2009)††		
					All Medicare§			Part D enrolled**				
		Number	% US*		Number	% US*		Number	% Medicare		Number	% US†
Total Enrollment		58 238 773	19.1		46 520 716	15.2		27 972 316	60.1		5 744 000	1.9
Gender												
	Female	31 512 082	54.1	Female	25 742 676	55.3	Female	16 517 856	59.1	Female	459 520	8
	Male	21 824 014	37.5	Male	20 778 040	44.7	Male	11 454 460	40.9	Male	5 280 800	92
	Unknown	4 902 677	8.4	—	—	—	—	—	—	—	—	—
Age												
	<1	2 006 749	3.4	—	—	—	—	—	—	—	—	—
	1–5	9 670 094	16.6	—	—	—	—	—	—	—	—	—
	6–12	9 516 371	16.3	—	—	—	—	—	—	—	—	—
	13–14	2 334 030	4.0	—	—	—	—	—	—	—	—	—
	15–18	4 837 569	8.3	<19	2658	0.0	—	—	—	—	—	—
	19–20	1 956 043	3.4	19-34	676 386	1.5	<65	5 550 293	19.8	—	—	—
	21–44	12 637 182	21.7	35-54	3 574 913	7.7	65-69	6 272 615	22.4	—	—	—
	45–64	5 639 547	9.7	55-64	3 501 360	7.5	70-74	5 355 331	19.1	<65	3 452 144	60.1
	65–74	2 011 317	3.5	65-74	20 606 076	44.3	75-79	4 205 434	15.03	≥65	2 291 856	39.9
	75–84	1 631 909	2.8	75-84	12 714 058	27.3	80-84	3 307 856	11.8	—	—	—
	85+	1 103 551	1.9	85+	5 445 265	11.7	85+	3 280 787	11.7	—	—	—
	Age group missing	4 894 411	8.4	—	—	—	—	—	—	—	—	—

(*Continued*)

Table 14.1 (*Continued*)

Characteristic		Medicaid (2008)‡		Medicare (2009)						Veterans Health Administration (2009)††		
				All Medicare§			Part D enrolled**					
		Number	% US*		Number	% US*		Number	% Medicare		Number	% US†
Race/ Ethnicity												
	Non-Hispanic White	22 135 196	38.0	White	38 589 288	83.0	White	22 438 315	80.2	Non-Hispanic White	4 554 992	79.3
	Black/African American	12 403 508	21.3	Black	4 727 383	10.2	Black	3 164 002	11.3	African American	649 072	11.3
	Hispanic or Latino	10 820 941	18.6	Hispanic	1 168 699	2.5	Hispanic	928 411	3.3	Hispanic or Latino	333 152	5.8
	Asian	1 446 157	2.5	Asian	898 809	1.9	Asian	686 657	2.5	Asian	86 160	1.5
	American Indian/Alaska Native	730 574	1.2	N. American Native	201 887	0.4	N. American Native	122 557	0.5	American Indian/Alaska Native	45 952	0.8
	Native Hawaiian/ Pacific Islander	465 114	0.8	Other	853 488	1.8	Other	570 323	2.0	Pacific Islander	-	-
	Hispanic/ Latino and one or more race	1 878 622	3.2	Unknown	81 162	0.2	Unknown	62 051	0.2	Other	74 672	1.3
	More than one race	147 061	0.3	—	—	—	—	—	—	—	—	—
	Not identified	8 211 600	14.1	—	—	—	—	—	—	—	—	—

*US Population estimates from (http: //www.census.gov/popest/states/NST-ann-est.html) N = 304 374 846 as of July 1, 2008.
†US Population estimates from (http: //www.census.gov/popest/states/NST-ann-est.html) N= 307 006 550 as of July 1, 2009.
‡FY 2008 Medicaid beneficiaries by gender (MSIS 2008 Table 13), age (MSIS 2008 Table 12), race/ ethnicity (MSIS 2008 Table 14).
§Medicare enrollment: hospital insurance and/or supplementary medical insurance enrollees, by demographic characteristics, as of July 1, 2009 (Table 2.3).
**Medicare Part D: type of coverage category for Part D enrollees, by demographic characteristics, as of December 2009 (Table 14.4).
††FY 2009 National Center for Veterans Analysis and Statistics, VA Benefits and Health Care Utilization. Veteran population as of Sept 30, 2009. http://www.va.gov/VETDATA/Pocket-Card/4X6_spring10_sharepoint.pdf.

of data requests, and providing technical assistance in the use of the data. Another valuable resource is a publication entitled *Pharmaceutical Benefits Under State Medical Assistance Programs*, which is published each year by the National Pharmaceutical Council (www.npcnow.org). The Kaiser Family Foundation also has a website with extensive detail on Medicaid state plans and coverage of specific services (http://medicaidbenefits.kff.org/index.jsp).

Pharmacoepidemiologic research has also been conducted since the early 1980s, using data obtained directly from individual states, including California,[7,8] Florida,[9] Iowa,[9] Missouri,[8] New Jersey,[9,10] New York,[11] Oregon,[8] and Tennessee.[12] Several commercial entities also make Medicaid data available (e.g., Thompson Reuters Healthcare).

Data structure of Medicaid databases

MAX files contain information on beneficiaries' demographics, enrollment, inpatient hospitalizations, outpatient physician visits, outpatient prescription drugs, outpatient laboratory and radiology studies, and stays in long-term care facilities (i.e., nursing homes, mental health hospitals). All records except for demographic and enrollment data originate from health care providers seeking reimbursement for goods and services. Such records are called claims. Hospitalization claims exclude many details of the hospital stay, including drugs administered. Outpatient laboratory and radiology claims report the type of test performed, but not results. Diagnoses are coded in the International Classification of Diseases, 9th Edition, Clinical Modification (ICD9-CM). Inpatient and outpatient procedures are coded in ICD-9-CM Procedure Codes, Common Procedure Terminology (CPT), Healthcare Common Procedure Coding System (HCPCS), or sometimes state-specific codes, depending on the state and file. ICD-10 will replace ICD-9 in the US on 1 October 2013. Drugs are coded according to the National Drug Code (NDC).

Medicaid data can be linked to Medicare to increase capture of care for dual enrollees.[13] Such a link can be very important, since Medicaid utilization records can fail to document a considerable proportion of care provided to dual enrollees.[14]

Medicare

Description of the Medicare program

Like Medicaid, Medicare was established in 1965. In contrast to Medicaid, Medicare is funded by the US federal government without state funds, and administered directly by CMS. Medicare provides health care coverage for nearly all legal residents of the US age 65 and above, some disabled people younger than 65, and people with end-stage renal disease or amyotrophic lateral sclerosis.

Medicare functions as a payer rather than as a direct provider of health care, and is made up of four separate parts: A, B, C, and D. All Parts of Medicare coverage require beneficiaries to pay deductibles; some stipulate cost sharing, where the beneficiary pays a percentage of costs, and premiums. Part A generally covers inpatient, short-term skilled nursing care, home health, and hospice care, and all Medicare beneficiaries are enrolled. Part B covers outpatient treatment and procedures, in addition to some drug therapies given in a physician's office or clinic, and is supplemental to Part A, with beneficiaries paying a monthly premium for coverage. Part C, also known as Medicare Advantage, is the Medicare managed care benefit. This was established in 1997 as part of the Balanced Budget Act.[15] Part C enrollees typically should not be included in pharmacoepidemiologic studies because claims for Part C enrollees are unavailable from CMS. Approximately 24% of Medicare beneficiaries are enrolled in Medicare Advantage plans, most of which are health maintenance organizations (HMOs);[16] this varies greatly with geography based on availability of HMOs.

Part D is the outpatient prescription drug coverage component of Medicare, which was instituted in 2006. It is administered by hundreds of private stand-alone prescription drug plans (PDPs), which supplement traditional fee-for-service Medicare (Parts A and B), as well as Medicare Advantage prescription drug (MA-PD) plans, which combine Parts A, B, and D in a managed care plan. PDP availability varies by state and, in 2010, a minimum of 41 stand-alone PDPs were available to beneficiaries in all states, with over 1500 total available in the US, in addition to various MA-PDs.[16] Each PDP

has its own formulary, with certain classes of drugs (e.g., benzodiazepines, barbiturates) excluded by law from all plans.

Deductibles for Medicare beneficiaries vary by type of service. For prescription drugs, in 2010, the standard Part D beneficiary paid the first $310 in prescription drug costs, with cost sharing for the patient being around 25% for the subsequent $2830 spent in a year. After that, the patient enters a coverage gap (also known as the "doughnut hole"), where they are responsible for 100% of drug costs ($3610 in 2010), until they reach the catastrophic coverage threshold.[16–18] Future legislation is projected to lower drugs costs for those in the coverage gap, while incrementally increasing the threshold for beneficiaries to enter the gap, with the hopes of closing the doughnut hole by 2020.[18,19] In addition, many states provide subsidy programs to low-income seniors to reduce out-of-pocket expenses for prescription drugs. For example, the Pennsylvania Pharmaceutical Assistance Contract for the Elderly (PACE) covers Part D premiums and provides reduced co-payments for low-income seniors ineligible for Medicaid.[20] In addition, with the expanded eligibility criteria for Medicaid, more Medicare recipients who find themselves in the doughnut hole will be able to rely on Medicaid to cover drug costs.

Characteristics of Medicare recipients

In 2010, Medicare covered 47 million Americans, with permanently disabled adults younger than 65 consisting of 8 million, or 17%, of Medicare recipients.[16] Baseline characteristics of the Medicare population are presented in Table 14.1. Females, whites, and the elderly are over-represented. Traditional fee-for-service (FFS) Medicare plans have a higher percentage of disabled beneficiaries under the age of 65, while Medicare Advantage plans have a higher percentage of Hispanic enrollees.[16]

Sources of Medicare data for research

CMS is the major source of Medicare data for researchers. ResDAC provides assistance to academic, government, and non-profit researchers in obtaining and using Medicare data.

Data structure of Medicare databases

Medicare data are available in numerous file types that are linkable to each other, as well as to Medicaid data for dually-enrolled beneficiaries. Some of these files contain information about enrollees, including date of birth, sex, address of residence, race, and, for decedents, death date. About half of the available files are claims-level standard analytic files (SAFs), which contain data based on claims submitted by providers. Institutional file types include: inpatient, outpatient, skilled nursing facility, hospice, and home health agency SAFs. Non-institutional data on physicians/ suppliers (carrier file), as well as durable medical equipment (DME), provide useful claims to researchers. Also available is the Medicare Provider and Analysis Review File (MedPAR), which includes inpatient and skilled nursing facility (SNF) final action claims. Unlike the SAFs, which are structured as one record per claim, the MedPAR file groups all claims for each inpatient or SNF stay into one line item. This makes it much easier to look at an entire hospitalization.

Of great interest to those engaged in pharmacoepidemiologic research is the prescription drug file. Prescription drug information in Medicare is available in the Part D Drug Event (PDE) file, which contains one record per dispensed prescription. Supplementary files are also available which provide information on Part D plans, pharmacies, drugs, and prescribers. Demographic information on Part D enrollees can be seen in Table 14.1.

The coding in Medicare files is similar to Medicaid. Diagnoses are coded in ICD-9-CM; inpatient and outpatient procedures are coded in a combination of ICD-9-CM Procedure Codes, CPT, and HCPCS codes, depending on the specific file. For example, MedPAR uses ICD-9-CM Procedures Codes only. Part D PDE files use NDCs to code individual drugs.

Department of Veterans Affairs Health Care System

Description of the Veterans Affairs

The Department of Veterans Affairs (VA) was established in 1930 as the Veterans Administration, when Congress authorized President Hoover to

"consolidate and coordinate Government activities affecting war veterans."[21,22] In March 1989, VA was elevated to a Cabinet level Department. The Department of Veterans Affairs' Veterans Health Administration (VHA) is one of the largest integrated health care systems in the United States, providing medical, surgical, and rehabilitative care to a diverse group of military Veterans who are mostly older, relatively sick, and often have multiple chronic medical and/or psychiatric conditions. VA also provides memorial services through the VA National Cemetery Administration, and financial and educational benefits through the Veterans Benefits Administration. In 2009, the VA healthcare system included 153 hospitals/ medical centers, over 1000 ambulatory care, mobile, independent, and community-based outpatient clinics, and 135 nursing homes.[22] The VA health care system consists of 21 regional integrated networks (Veteran Integrated Service Networks or VISNs).

In contrast to the Medicaid and Medicare programs, which act exclusively as payers, the VA health care system is primarily a direct provider of health care services, funded by the US Government. Veterans who served in the military, as well as active duty Reservists and National Guard, are potentially eligible for VA health care.[21] Veteran eligibility requirements are variable and depend on length of military service, type of discharge, and time of service.[23] There are eight priority groupings used to rank eligibility status of veterans. The priority groups are defined by many factors, some of which include: various service-connected disabilities, theaters of war, and socioeconomic status. While veterans receiving health care are not required to pay premiums for coverage, some are charged co-payments for certain medical services and outpatient prescriptions, depending on their priority group.[22] With minor exceptions, only prescriptions written by VA prescribers and filled in a VA pharmacy are covered by the VA's comprehensive medical care plan.

Characteristics of Veterans Affairs population

In 2009, approximately 5.7 million veterans were treated in the VA health care system, with over 4.5 million receiving prescriptions.[24] Table 14.1 presents demographic characteristics of VA patients, which differ from those of the general adult population in terms of age and gender. Almost 40% of veterans were over 65 years of age in 2009, and 92% of Veteran beneficiaries are male.

Sources of Veterans Affairs data for research

Access to VA data is limited to researchers employed by the VA (including academics with a VA appointment, either with or without pay) and their collaborators.

Data structure of Veterans Affairs databases

Pharmacy data systems record drugs dispensed to inpatients in VA hospitals as well as outpatient prescriptions dispensed by its outpatient pharmacies, 80% of which are dispensed through the mail via VA's Consolidated Mail-order Pharmacies (CMOPs). Drugs are recorded and classified according to the VA Drug Classification System (http://www.pbm.va.gov/NationalFormulary.aspx), which is similar to the American Hospital Formulary Service categorization of pharmaceutical products. Clinical databases contain data on inpatient and outpatient encounters, including admissions, discharges, transfers, clinic visits, prescription orders, laboratory, radiology, surgery, and administrative services. Diagnoses are recorded in ICD-9, while ICD-10 will be used beginning in 2013. Death data are obtained from various sources, with cross-checks with the Social Security Administration Death Master File.[25] In addition to the VA administrative databases, the VA has several disease-specific registries that are used for patient care and research. Examples of conditions with existing registries include cancer, diabetes, HIV, and hepatitis C. The VHA also maintains its own adverse drug event reporting system (also see Chapter 10), which currently contains over 200 000 reports related to drugs or vaccines.

Strengths

Population size and length of follow-up

An important strength of these databases is their large size: 58.2 million people in Medicaid, 46.5

million in Medicare, and 5.7 million in the VA. These numbers are expected to grow with the extension of Medicaid benefits, the aging of the US population, and because of the current number of active duty military, Reservists, and National Guard serving in current conflicts in the Middle East. Because Medicaid and Medicare files are linkable to each other, beneficiaries can be followed across years, even if Medicaid enrollees join Medicare. With Medicare and the VA, patients who enter the system often remain for many years, permitting long-term follow-up. In contrast, there is much greater turnover in Medicaid, limiting the ability to study long-term effects of drugs.

Accuracy of pharmacy claims

Another notable strength of these data is that pharmacy claims record what was dispensed by the pharmacy, which is one step closer to ingestion than what was prescribed, as recorded in medical record databases (see Chapter 15). Further, outpatient prescription claims accurately record the date, drug, and quantity dispensed by the pharmacy. This is because this information determines the payment provided to the pharmacy, and is subject to regulatory audit. Thus, a validation study of Medicaid pharmacy claims in the early 1980s found them to be highly accurate.[26] Traditionally, because of low co-payments, patients have had a strong financial incentive to use these programs to purchase their drugs instead of paying for them out-of-pocket. This served to increase the completeness of pharmacy claims data. However, with the recent offering of low priced (e.g., $4) generic prescriptions by many retail pharmacies,[27] patients and pharmacists may have little financial incentive to report low-cost prescriptions to the patient's drug benefit plan, and thus some prescriptions may not be recorded in US claims databases. Similarly, Medicare beneficiaries in the coverage gap may have little financial incentive to have their prescriptions submitted to their Medicare prescription drug program, particularly if they do not expect to exceed the coverage gap. Further, patients may not take all of the medicine dispensed as directed. However, for chronically administered drugs, dispensing records have been found to accurately reflect cumulative exposure and gaps in medication supply when compared with electronic medication containers that record when the bottle was opened,[28] which are currently considered the best available means of measuring medication ingestion in community-dwelling patients (see Chapter 42).

In the VA, the outpatient database tracks prescription medications and non-prescription medications obtained through the VA, as well as specific medical supplies. As is the case with medications obtained through Medicaid services, the low co-payment, and in some instances no co-payment, associated with VA prescriptions has resulted in a strong financial incentive for Veterans to obtain their outpatient prescriptions through the VA. An additional advantage of the VA data is the recording of drugs dispensed by the pharmacy to patients hospitalized in VA hospitals. However, certain medications (e.g., those obtained from floor stock and medications administered acutely in acute care areas), though recorded in the electronic medical records, are not accessible in the prescription databases. This unavailability can be attributed to the way in which these medications are recorded and the location where the medication is provided.

Validity of procedure claims

For Medicare and Medicaid, claims with codes for clinical procedures determine the amount of money paid to the health care provider. Therefore, procedure records are audited to detect fraud, and would thus be expected to be highly accurate with regard to performance of that procedure. Wysowski *et al*. performed medical record validation in a Medicaid study that used presence of a surgical procedure code as part of an algorithm to identify cases of hip fracture.[29] They found 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.

Over-representation of underserved populations

Another potential strength of Medicaid is over-representation of traditionally underrepresented groups. Medicaid has substantially greater numbers

of pregnant women, young children, and African-Americans than other data sets. Seniors, who bear the greatest burden of medication-related adverse events, make up 86% of the Medicare population, 40% of the VA population, and 8% of Medicaid. This is particularly important given that seniors are underrepresented in US commercial insurance databases (see Chapter 13).

Ability to validate outcomes

Examining the validity of diagnosis codes frequently requires review of clinical records. Fortunately, a mechanism exists to obtain inpatient hospital and emergency department records corresponding to Medicare and Medicaid claims.[30] Using this mechanism, researchers have been able to obtain approximately 70–75% of inpatient hospital and emergency department records. The cost of obtaining medical records, however, is not inexpensive (approximately $150/record in 2010). To our knowledge, this mechanism has not been tried for obtaining outpatient medical or dental records.

One of the strengths of the VA data is the mechanism for obtaining primary inpatient and outpatient medical records to validate outcomes. Records can be obtained electronically from the local healthcare system, where the medical care information is current and complete.

Ability to link to external data

In addition to the capability to link to each other, Medicaid and Medicare data have also been linked to sources of mortality data such as the Social Security Administration Death Master File,[31] National Death Index,[32] and state vital statistics registries.[33] VA data can also be linked to Medicaid and Medicare and National Death Index data. Medicare data have also been linked to data from state pharmaceutical assistance programs for the elderly to identify outcomes in enrollees of the state programs.[34] Linkage to birth certificate data has been performed for studies of the effects of fetal exposure to medications and for evaluations of the effects on newborns of policies that affect prenatal care.[35] Drivers' license data and police reports of injurious crashes have also been linked to Medicaid data.[36] Studies using registries (see Chapter 21) have linked to Medicare, Medicaid, and VA data to identify outcomes.[37–39] Medicare data can be linked to the Medicare Current Beneficiary Survey (MCBS)[40] to obtain additional information about the subgroup of subjects who participated in the MCBS. Data from Medicaid and Medicare enrollees admitted to a nursing home can be linked to the nursing home Minimum Data Set to obtain additional information, including measures of physical, psychological, and psychosocial functioning in these enrollees.[41] External data obtained on a subset of such linkages can be used to adjust for factors not recorded in the parent database[42] (see also Chapter 47). Of note, CMS does not permit data linkage using beneficiary name and address; researchers can request the encrypted Bene_ID, a unique identifying number assigned to each beneficiary in Medicaid and Medicare, to accurately link files spanning data sources, states, and years.

Weaknesses

Non-representativeness

As discussed above, each of the programs described in this chapter is unrepresentative of the general population in a different way. Non-representativeness may be an important limitation for descriptive studies seeking to describe the overall population and its health care utilization. For example, newborn deliveries account for 40% of hospital admissions within Medicaid beneficiaries, but only 16% of admissions in the non-Medicaid US population.[43] However, for etiologic studies, generalizability is compromised only for biologic relationships that vary by factors that differ between the studied and general populations. For example, Medicaid studies evaluating the gastrointestinal side effects of nonsteroidal anti-inflammatory drugs have produced similar results to those performed in other populations.[44] On the other hand, studying groups with a high degree of comorbidity (e.g., Medicaid and VA beneficiaries) may be advantageous in some circumstances, such as when drug effects are more readily discernable in high-risk groups. For example, cisapride was found to be

associated with an elevated risk of ventricular arrhythmia in a Medicaid population[45] but not in two general, unselected populations.[46] Further, this homogeneity can sometimes assist in controlling for confounding, for example controlling for socioeconomic status.

Unavailable information

Administrative and clinical data like those described here often lack information on many potentially important confounding factors, such as smoking, exercise, diet, environmental exposures, illegal drug use, alcohol use, occupation, family history, and use of many non-prescription drugs. Some of these factors can be obtained in a subset of patients by reviewing primary medical records to the degree that these factors are recorded in medical records. Linkages with external data sources, such as those described above, can also provide additional information for all or a subset of subjects.

Limitations in prescription coverage

Only drugs covered by Medicare and Medicaid can be studied using the encounter data. A number of drug categories are generally not covered by the Medicare and Medicaid prescription drug benefits, such as agents for fertility, weight loss, hair growth, cosmetic effect, and over-the-counter smoking cessation. Coverage of non-prescription drugs varies by state in Medicaid, and is not included in Medicare Part D. Therefore, these agents cannot be studied using these data. Prior approval is also required before reimbursing for certain drugs, such as human growth hormone, non-sedating antihistamines, and more expensive nonsteroidal anti-inflammatory agents.

In Medicare, the Part D program is carried out by over a thousand private pharmacy benefit plans, each offering a selection of formularies and cost-sharing options. This leads to inconsistencies in drug availability across plans, which may limit which products beneficiaries can access. The coverage inconsistencies and plan options have also been shown to be confusing to most seniors.[47] With beneficiaries paying for 100% of drug costs in the current coverage gap, they lack incentive to file their claim through Medicare, and a decrease in prescription days has been noted.[48] This would lead to missing data for the beneficiaries who do not expect to reach catastrophic coverage, and may stop filing claims when they reach the doughnut hole. Researchers may also experience gaps in prescription records for beneficiaries who are paying out-of-pocket for prescriptions because they do not anticipate exceeding the plan's deductible.

In Medicaid, the coverage of injectable drugs and adult vaccines varies by state, although coverage for many childhood vaccines is required by federal law. Whether injectable drugs are recorded as prescription encounters or other types of encounters also varies by state. Medicare covers the cost of certain adult vaccinations and their administration (e.g., influenza, pneumococcal pneumonia, hepatitis B).

All prescriptions written by VA prescribers and filled in a VA pharmacy are covered by the VA's comprehensive medical care plan. The VA National Formulary lists all products (drugs and supplies) covered under the VA medical care plan. Any agent not on the Formulary and approved through a non-formulary mechanism is also covered. Only drugs dispensed and released by the VA can be studied using the VA prescription databases and corresponding medical record data. The greatest limitation to the prescription database occurs with the inpatient dispensing of agents in acute care settings, as described in the section Accuracy of Pharmacy Claims, above.

Eligibility and data limitations

An important potential challenge for all US claims databases is the availability of low cost prescriptions, which may reduce or eliminate incentives to submit prescription claims for payment. This may be less problematic in the low-resource populations using Medicaid and VA health care, but there are no data to confirm that.

When using Medicaid data, researchers may also experience gaps in beneficiary records due to periods of ineligibility. Since Medicaid is known to have frequent turnover in their beneficiaries, researchers must develop strategies to deal with this limitation. One approach to this is to use the Medicaid eligibility files. However, these files may

not be completely accurate. Another approach to reducing this potential problem is to restrict consideration to time periods in which Medicaid encounters are present within some specified period (e.g., 6 months) both before and after the person-time under study,[25,26,33,37–40] or there is evidence of death. There are no data to determine which of these approaches is preferable.

In both Medicaid and Medicare, there is a data flag to indicate whether or not a beneficiary has been enrolled in managed care. This information can be used to exclude or label beneficiaries, who may have eligibility files in the database, but for subjects enrolled in capitated plans, it is uncertain whether encounter-level information such as hospitalizations and physician visits will be recorded in the encounter files. Since 1999, Medicaid has required states to provide CMS with encounter data for individuals enrolled in capitated plans. However, despite this requirement, encounter data for those enrolled in capitated plans appears to be incomplete in at least some states. The problem of missing encounter data for persons enrolled in capitated plans can be avoided by excluding person-time during which the individual was enrolled in a capitated plan.

In Medicare and the VA, in contrast, turnover is low, as beneficiaries of both programs often remain enrolled once they are deemed eligible to receive benefits.

Data validity/access to medical records

The validity of data on exposures, outcomes, and covariates is a major consideration when using any pre-existing data for research. It is important to keep in mind that these data were generated as a by-product of providing health care or administering health care benefits rather than for research purposes. This is true for all administrative and medical record databases. As a result, researchers need to consider whether a given research question can be addressed using pre-existing data.

Our experience suggests that in each study, with few exceptions, investigators should obtain medical records in at least a sample of outcomes to confirm the validity of the encounter diagnoses, characterize the severity of the disease, and obtain information on potential confounding variables not found in the encounter data. One potential exception is studies of outcomes for which encounter diagnoses have previously been found to be sufficiently valid. Another potential exception is studies using a procedure or a prescription for a drug as the outcome of interest.[49]

Out-of-plan care

The lack of completeness of VA data with regard to health care obtained outside the VHA is an important issue. Veteran patients can voluntarily go to any hospital for care. Moreover, for emergency and urgent conditions, Veteran patients are taken to the nearest hospital for care. Due to this, many acute conditions such as myocardial infarction, stroke, and severe hypoglycemic events are not captured as inpatient events and these important outcomes can be missed. For patients under the age of 65, these missing events must be taken into consideration for any study. For studies of VA enrollees aged 65 and above, out-of-plan medical care can often be identified by linking VA data to Medicare data. Because Medicaid and Medicare reimburse healthcare providers in the private sector, out-of-plan care is less likely to be an issue in those plans than in the VA.

Particular applications

Many methodologic and applied studies have been performed using data from the Medicaid, Medicare, and VA systems. A few illustrative examples are presented here.

Methodologic studies

Stürmer and colleagues used Medicaid data from New Jersey to compare three approaches to adjusting for measured confounding variables: conventional adjustment using the observed variables, adjustment using propensity scores (see Chapter 47), and adjustment using disease risk scores (see Chapter 47).[50] They found that the three methods all produced similar results.

Hennessy and colleagues performed descriptive analyses to assess the integrity of data that were provided by a commercial vendor for six Medicaid programs.[14] They found that prescription encounter records appeared to be intermittently missing in some states and that there was no valid marker for inpatient hospitalizations for some states. In addition, hospitalizations in those aged 65 years and above appeared to be missing to varying degrees in all states, presumably because Medicare was the primary payer for such hospitalizations. Mismatches between diagnostic and demographic information (e.g., female disorders in males) were rare. The authors recommended that whenever possible, descriptive analyses of the underlying administrative data be used to identify potentially important data anomalies.[14]

McKenzie and colleagues examined the validity of Medicaid pharmacy encounter records to estimate drug use in elderly nursing home residents.[51] They found good agreement between Medicaid encounter records and nursing home records for presence or absence of drug ingestion (positive and negative predictive values >85%), and that doses recorded using the two databases correlated well (correlation coefficients from 0.66 to 0.97).[51]

Schneeweiss and colleagues used Medicare data linked to a state pharmaceutical assistance program for the elderly to assess high-dimensional propensity scores (see Chapter 47) as an approach to controlling for measured confounding factors.[34] They found that high-dimensional propensity scores resulted in effect estimates closer to those produced by randomized trials than did conventional adjustment by predefined covariates.

Several studies have been conducted with VA data to assess the validity of ICD-9 codes to identify specific conditions. For example, Petersen *et al*. assessed the predictive value of myocardial infarction (MI), coronary bypass graft surgery, cardiac catheterization, and angioplasty. The positive predictive values of claims ranged from 90 to 100%.[52]

Applied pharmacoepidemiologic studies

Roumie *et al*. used Tennessee Medicaid data to examine the association between different nonsteroidal anti-inflammatory drugs and stroke, myocardial infarction, and cardiovascular death. They found that current use of rofecoxib, valdecoxib, and indomethacin was associated with an increased risk of cardiovascular events in those without preexisting cardiovascular disease.[53]

Ray and colleagues used Tennessee Medicaid data to examine the association between antipsychotic drugs and risk of sudden cardiac death.[33] They found that at doses of greater than 100 mg/d of chlorpromazine equivalents, the rate ratio for use of any antipsychotic drug was 2.39 (95% confidence interval [CI], 1.77 to 3.22).[33] Hennessy and colleagues used Medicaid data from three states to study the risk of a composite outcome of sudden death or ventricular arrhythmia in persons with schizophrenia who received antipsychotics.[31] The primary comparison was thioridazine versus haloperidol. They found no overall difference in the rate of the composite outcome, although thioridazine had a higher risk of the composite outcome at doses of 600 mg/d or greater in chlorpromazine equivalents (rate ratio 2.6, 95% CI 1.0 to 6.6). A dose–response relationship was evident for thioridazine but not for haloperidol.[31]

Patrick *et al*. used data from Medicare and the Pennsylvania state pharmaceutical assistance plan to examine the relationship between adherence to bisphosphonates and fracture risk.[54] They found that good adherence (defined as 80–100% of days covered; see Chapter 42) was associated with a 22% reduction in overall fracture rate compared to worse adherence.

Lambert *et al*. evaluated the risk of new-onset diabetes in Veterans exposed to specific antipsychotic agents.[55] They found that the second-generation antipsychotic agents studied were associated with a higher risk of diabetes than first generation agents.

The future

Because Medicaid has covered drugs and medical care for decades, its data have a long history of use in pharmacoepidemiology. Since Medicaid data have become available from CMS with assistance

provided by ResDAC, use of Medicaid data for research has increased even more, and will probably continue to do so. With the anticipated expansion of Medicaid, the size of the population available should increase further, making these data even more valuable.

The new Medicare drug benefit should be an enormously valuable resource for pharmacoepidemiology, and this is beginning to occur. We are hopeful that CMS will continue to permit researchers to obtain medical records to validate outcomes identified in Medicaid and Medicare data. In the future, perhaps it will be possible to access Medicare patients directly, obtaining supplemental historical and even biologic information.

VA databases have existed for well over a decade but continue to evolve to include more detailed clinical information. In the future, more real-time databases will be available, as will newer VA databases, many of which are now in the pilot phase. These newer databases should allow for more timely evaluation and the ability to better evaluate population trends at the time of care. Some of the projects that are currently in an early or pilot phase are newer methods for creating common prevalent cohorts available for medical chart validations, automated text-based validation tools, and increased use of VA databases for comparative effectiveness research (see Chapter 32).

Also, databases that are currently being compiled by the Department of Defense (DoD) could prove very helpful in the examination of vaccines, as well as other medications. These DoD databases hold great potential for pharmacoepidemiologic research studies.

Finally, the potential use of US Government health care data as part of FDA's Sentinel initiative for active surveillance (see Chapter 30) holds great promise.

Acknowledgments

We thank Gerri Barosso, RD, MPH, MS for commenting on a draft of this chapter. Ms Barosso's work is supported by CMS contract HHSM 500200500027I to the University of Minnesota School of Public Health.

References

1. Klees B, Wolfe C, Curtis C. *Medicaid Program Description and Legislative History, 2010*. 7500 Security Blvd., Baltimore, MD 21244, Office of the Actuary, Centers for Medicare & Medicaid Services. U.S. Social Security Administration, Office of Retirement and Disability Policy. Annual Statistical Supplement, 2010.
2. Kaiser Family Foundation. *Medicaid: Overview and Impact of New Regulations*. Washington, DC: The Henry J. Kaiser Family Foundation, 2008. Report No. #7739.
3. Heberlein M, Guyer J, Rudowitz R. *Financing New Medicaid Coverage Under Health Reform: The Role of the Federal Government and States*. Kaiser Commission on Medicaid and the Uninsured, Kaiser Family Foundation, 2010. Report No. #8072.
4. National Pharmaceutical Council. *Pharmaceutical Benefits Under State Medical Assistance Programs, 2007*. Reston, VA: National Pharmaceutical Council, 2007.
5. Briesacher BA, Soumerai SB, Field TS, Fouayzi H, Gurwitz JH. Medicare part D's exclusion of benzodiazepines and fracture risk in nursing homes. *Arch Intern Med* 2010; **170**: 693–8.
6. Kaiser Family Foundation. *Medicaid and Long-Term Care Services and Support*. Washington, DC: The Kaiser Commission on Medicaid and the Uninsured, 2009. Report No. #2186–06.
7. Menzin J, Boulanger L, Friedman M, Mackell J, Lloyd JR. Treatment adherence associated with conventional and atypical antipsychotics in a large state Medicaid program. *Psychiatr Serv* 2003; **54**: 719–23.
8. Morrato E, Nicol G, Maahs D, Druss B, Hartung D, Valuck R, *et al*. Metabolic screening in children receiving antipsychotic drug treatment. *Arch Pediatr Adolesc Med* 2010; **164**: 344–51.
9. Faught R, Weiner J, Guerin A, Cunnington M, Duh M. Impact of nonadherence to antiepileptic drugs on health care utilization and costs: Findings from the RANSOM study. *Epilepsia* 2009; **50**: 501–9.
10. Wang PS, Walker AM, Tsuang MT, Orav EJ, Glynn RJ, Levin R, *et al*. Dopamine antagonists and the development of breast cancer. *Arch Gen Psychiatry* 2002; **59**: 1147–54.
11. Laine C, Zhang D, Hauck WW, Turner BJ. HIV-1 RNA viral load monitoring in HIV-infected drug users on antiretroviral therapy: relationship with outpatient

care patterns. *J Acquir Immune Defic Syndr* 2002; **29**: 270–4.
12. Ray WA, Meredith S, Thapa PB, Hall K, Murray KT. Cyclic antidepressants and the risk of sudden cardiac death. *Clin Pharmacol Ther* 2004; **75**: 234–41.
13. Wang PS, Levin R, Zhao SZ, Avorn J. Urinary antispasmodic use and the risks of ventricular arrhythmia and sudden death in older patients. *J Am Geriatr Soc* 2002; **50**: 117–24.
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**: 103–11.
15. Klees B, Wolfe C, Curtis C. *Brief Summaries of Medicare and Medicaid*, 2010: 1–27. https://www.cms.gov/MedicareProgramRatesStats/02_SummaryMedicareMedicaid.asp (accessed June 2011).
16. Cubanski J, Huang J, Damico A, Jacobson G, Neuman T. *Medicare Chartbook*, 4th edn. Kaiser Family Foundation, 2010. Report No. 8103.
17. DePue RJ, Jr, Stubbings J, Baker DC. Medicare Part D policy update and implications for 2010. *Am J Manag Care* 2010; **16**: e117–e120.
18. Kaiser Family Foundation. *Explaining Health Care Reform: Key Changes to the Medicare Part D Drug Benefit Coverage Gap*, 2010. Report No. #8059.
19. Shrank WH, Choudhry NK. Time to fill the doughnuts—health care reform and Medicare Part D. *N Engl J Med* 2011; **364**: 598–601.
20. Pennsylvania Department of Aging. *Prescription Assistance, PACE and PACENET*, 2011 http://www.portal.state.pa.us/portal/server.pt/community/pace_pacenet/17944 (accessed June 2011).
21. US Department of Veterans Affairs. *VA History*. 810 Vermont Avenue, NW-Washington, DC 20420, US Department of Veterans Affairs, 2010.
22. Panangala S. *Veterans Medical Care: FY2011 Appropriations*. Washington, DC, Congressional Research Service. CRS Report for Congress, 2010.
23. US Department of Veterans Affairs. *VA Health Care Eligibility and Enrollment*, 2010. http://www4.va.gov/healtheligibility/eligibility/PriorityGroupsAll.asp (accessed June 2011).
24. US Department of Veterans Affairs. National Center for Veterans Analysis and Statistics. *Demographics*. http://www.va.gov/VETDATA/Demographics/Demographics.asp (accessed June 2011).
25. Sohn M, Arnold N, Maynard C, Hynes D. Accuracy and completeness of mortality data in the Department of Veterans Affairs. *Popul Health Metr* 2006; **4**: 2–8.
26. Lessler JT, Harris BSH. *Medicaid Data as a Source for Postmarketing Surveillance Information*. Research Triangle Park, NC: Research Triangle Institute, 1984.
27. Choudhry NK, Shrank WH. Four-dollar generics—increased accessibility, impaired quality assurance. *N Engl J Med* 2010; **363**: 1885–7.
28. Choo PW, Rand CS, Inui TS, Lee ML, Cain E, Cordeiro-Breault M, *et al*. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. *Med Care* 1999; **37**: 846–57.
29. Wysowski DK, Baum C. The validity of Medicaid diagnoses of hip fracture. *Am J Public Health* 1993; **83**: 770.
30. Hennessy S, Leonard CE, Bilker WB. Researchers and HIPAA. *Epidemiology* 2007; **18**: 518.
31. Hennessy S, Bilker W, Knauss J, Margolis D, Kimmel S, Reynolds R, *et al*. Cardiac arrest and ventricular arrhythmia in patients taking antipsychotic drugs: cohort study using administrative data. *BMJ* 2002; **325**: 1070.
32. Staffa JA, Jones JK, Gable CB, Verspeelt JP, Amery WK. Risk of selected serious cardiac events among new users of antihistamines. *Clin Ther* 1995; **17**: 1062–77.
33. Ray WA, Meredith S, Thapa PB, Meador KG, Hall K, Murray KT. Antipsychotics and the risk of sudden cardiac death. *Arch Gen Psychiatry* 2001; **58**: 1161–7.
34. 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**: 512–22.
35. Ray WA, Gigante J, Mitchel EF, Jr, Hickson GB. Perinatal outcomes following implementation of TennCare. *Chest* 1998; **279**: 314–6.
36. 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.
37. Wang PS, Walker AM, Tsuang MT, Orav EJ, Glynn RJ, Levin R, *et al*. Dopamine antagonists and the development of breast cancer. *Arch Gen Psychiatry* 2002; **59**: 1147–54.
38. Hammill B, Hernandez A, Peterson E, Fonarow G, Schulman K, Curtus L. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. *Am Heart J* 2009; **157**: 995–1000.
39. Bradley C, Given C, Luo Z, Roberts C, Copeland G, Virnig B. Medicaid, Medicare, and the Michigan Tumor Registry: a linkage strategy. *Med Decis Making* 2007; **27**: 352–63.

40. DiMartino LD, Shea AM, Hernandez AF, Curtis LH. Use of guideline-recommended therapies for heart failure in the Medicare population. *Clin Cardiol* 2010; **33**: 400–5.
41. Crystal S, Gaboda D, Lucas J, Gerhard T, Chakravarty S. Assessing medication exposures and outcomes in the frail elderly: assessing research challenges in nursing home pharmacotherapy. *Med Care* 2010; **48** (6 Suppl): S23–S31.
42. 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** (Suppl 2): S158–S165.
43. National Center for Health Statistics. *National Hospital Discharge Survey 2001*. Hyattsville, MD: National Center for Health Statistics, 2003.
44. Henry D, Lim LL, Garcia Rodriguez LA, Perez GS, Carson JL, Griffin M, *et al*. Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta-analysis. *BMJ* 1996; **312**: 1563–6.
45. Hennessy S, Leonard C, Newcomb C, Kimmel S, Bilker W. Cisapride and ventricular arrhythmia. *Br J Clin Pharmacol* 2008; **66**: 375–85.
46. Walker AM, Szneke P, Weatherby LB, Dicker LW, Lanza LL, Loughlin JE, *et al*. The risk of serious cardiac arrhythmias among cisapride users in the United Kingdom and Canada. *Am J Med* 1999; **107**: 356–62.
47. Polinski JM, Bhandari A, Saya UY, Schneeweiss S, Shrank WH. Medicare beneficiaries' knowledge of and choices regarding Part D, 2005 to the present. *J Am Geriatr Soc* 2010; **58**: 950–66.
48. Polinski JM, Kilabuk E, Schneeweiss S, Brennan T, Shrank WH. Changes in drug use and out-of-pocket costs associated with Medicare Part D implementation: a systematic review. *J Am Geriatr Soc* 2010; **58**: 1764–79.
49. Avorn J, Gurwitz JH, Bohn RL, Mogun H, Monane M, Walker A. Increased incidence of levodopa therapy following metoclopramide use. *Chest* 1995; **274**: 1780–2.
50. Stürmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. *Am J Epidemiol* 2005; **161**: 891–8.
51. McKenzie DA, Semradek J, McFarland BH, Mullooly JP, McCamant LE. The validity of medicaid pharmacy claims for estimating drug use among elderly nursing home residents: The Oregon experience. *J Clin Epidemiol* 2000; **53**: 1248–57.
52. Petersen L, Wright S, Normand S, Daley J. Positive predictive value of the diagnosis of acture myocardial infarction in an administrative database. *J Gen Intern Med* 1999; **14**: 555–8.
53. Roumie CL, Choma NN, Kaltenbach L, Mitchel EF, Jr., Arbogast PG, Griffin MR. Non-aspirin NSAIDs, cyclooxygenase-2 inhibitors and risk for cardiovascular events-stroke, acute myocardial infarction, and death from coronary heart disease. *Pharmacoepidemiol Drug Saf* 2009; **18**: 1053–63.
54. Patrick AR, Brookhart MA, Losina E, Schousboe JT, Cadarette SM, Mogun H, *et al*. The complex relation between bisphosphonate adherence and fracture reduction. *J Clin Endocrinol Metab* 2010; **95**: 3251–9.
55. Lambert B, Cunningham F, Dalack G, Hur K, Miller D. Diabetes risk associated with use of olanzapine, quetiapine and risperidone in VA patients with schizophrenia. *Am J Epidemiol* 2006; **164**: 672–81.
56. Centers for Medicare and Medicaid Services. *Medicare and Medicaid Statistical Supplement*, 2010. http://www.cms.gov/MedicareMedicaidStatSupp/ (accessed 2011).
57. Centers for Medicare and Medicaid Services. *MSIS Tables*, 2010. http://www.cms.gov/MedicaidDataSourcesGenInfo/MSIS/list.asp (accessed 2011).
58. US Census Bureau. *National and State Population Estimates*.http://www.census.gov/popest/states/NST-annest.html (accessed 2011).
59. US Department of Veterans Affairs. National Center for Veterans Analysis and Statistics. *VA Benefits and Health Care Utilization*, 2010. http://www.va.gov/VETDATA/Pocket-Card/4X6_spring10_sharepoint.pdf (accessed June 2011).

CHAPTER 15
Medical Record Databases

Alexis Ogdie,[1] Sinéad M. Langan,[2] John Parkinson,[3] Hassy Dattani,[4] Karel Kostev,[5] and Joel M. Gelfand[2,6]

[1] Division of Rheumatology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[2] Department of Epidemicology and Public Health, London. School of Hygiene and Tropical Medicine, London, UK

[3] General Practice Research Database (GPRD), Medicines and Healthcare products Regulatory Agency (MHRA), London, UK

[4] Cegedim Strategic Data Medical Research Ltd. London, UK

[5] Centre of Excellence Patient Data, IMS Health GmbH & Co OHG, Frankfurt/Main, Germany

[6] Department of Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

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, and those that serve as the patient's medical record and therefore are a primary means by which physicians track health information about their patients. Administrative databases (e.g., commercial insurance databases, described in Chapter 13, and government databases like Medicaid, described in Chapter 14) are maintained for the purpose of billing or otherwise administering care, rather than for the actual provision of patient care. As a result, while the number of patients within these databases is frequently large, administrative databases often fail to capture important information about the patient, such as smoking status, alcohol use, and body mass index, which is generally recorded in medical record databases.[1] Further, diagnosis codes entered for administrative rather than clinical purposes may not reflect the clinician's view of the true clinical state of the patient. In contrast, information from medical record databases may be more likely to reflect the patient's true clinical state, because the data are collected for patient care. Databases derived from systems where many patients receive out-of-system care (such as the Veterans Affairs system, described in Chapter 14) may suffer from incompleteness in the recording of diagnoses and treatments, unlike closed systems such as some insurance or primary care medical record databases, in which all of the patients' health-care experiences are ostensibly captured.

Medical record databases are longitudinal patient record databases that are used by health care providers in caring for their patients, and anonymized for the purpose of research. However, the completeness of primary care medical record databases in recording outside care should be evaluated rather than assumed. Further, data from medical record databases still require careful study to assess the validity of the exposure and diagnosis data. In addition, there may be a high proportion of missing or absent data for some variables of interest, such as smoking, alcohol use, and occupation. In this chapter, we focus on primary care medical record databases, electronic patient records taken from interactions with primary care givers, and in some cases specialists based in the outpatient setting, which include information on past and current medical problems and therapies including prescriptions and other modalities.

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

The medical record databases highlighted in this chapter include the General Practice Research Database (GPRD), The Health Improvement Network (THIN), and the Intercontinental Marketing Services (IMS) Disease Analyzer (previously known as Mediplus). All three are available for licensing by investigators in industry, government, and academia. Other more localized electronic medical record databases exist but are less likely to be useful by themselves in pharmacoepidemiologic research, either because of incomplete data collection, small populations, or other reasons. Therefore, they are beyond the scope of this chapter. GPRD and THIN are derived from the medical records of patients within the United Kingdom (UK), whereas Disease Analyzer contains records from France, Germany, and UK. These databases have been used by epidemiologists, and in particular pharmacoepidemiologists, resulting in hundreds of published articles. While there are many similarities among the databases, there are also important differences which we describe in more detail (Table 15.1).

The UK has advantages for obtaining electronic medical records data as a "gatekeeper" system exists in which all patients must be registered with general practitioners (GP) to get "free at the point of care" National Health Service (NHS) treatment.[7] Almost the entire population in the UK is registered with a GP[8] and most GPs have computerized medical records. GPs are informed of all medical events including secondary and tertiary care. The UK was the setting of the first of these databases, GPRD, established in 1987 as a tool for conducting public health research. Since then, several more medical record databases have been developed for research purposes. The IMS Disease Analyzer, while including patients from the UK, also includes patients from Germany and France. Both Germany and France have universal health care, but patients frequently have additional private insurance. Unlike the UK, medical care in Germany and France is not necessarily driven by the primary care physicians, as patients may see a specialist without first visiting a GP depending on their insurance coverage.[9–12]

The GPRD, initially called Value Added Medical Products (VAMP) Research Databank, provided GPs with software that enabled them to contribute anonymized patient information to a central database.[13] In order to participate, GPs agreed to undergo training in data entry and provide a copy of the entire anonymous medical record to VAMP including photocopied (but de-identified) letters from specialists and hospitals. Since inception, there have been many changes in management of the database. The Medicines and Healthcare products Regulatory Agency (MHRA) has been the single vendor since 2003. Since 1995, GPRD has used proprietary software known as Vision for obtaining patient information.

The Health Improvement Network began collecting data in 2002, also using Vision. THIN is a collaboration between Cegedim (owner of In Practice Systems, INPS) and Cegedim Strategic Data (Medical Research Company,) Ltd[14] (formerly EPIC). INPS disperses and supports Vision software whereas EPIC manages data extraction and database use. GPRD and THIN collect similar information from representative populations within the UK.[15] Some practices participate in both THIN and GPRD.[15] The most recent unpublished estimate from 2009 suggested that roughly 30% (GPRD internal data, not a validated figure) contribute to both databases, although this is changing as new practices are added to one or both databases. It is important to recognize that pooling results from both databases would be problematic given that the datasets are not independent. Lewis *et al*. used GPRD practices and non-GPRD practices within THIN to explore outcomes including colorectal cancer, stroke, peptic ulcer disease, myocardial infarction, and relevant exposures among GPRD practices and non-GPRD practices within THIN. Matched case–control studies were performed for each of the outcomes; the investigators found that the GPRD and non-GPRD practices had similar associations and magnitude of the associations.[15] The IMS Disease Analyzer differs from GPRD and THIN in that data are collected from three countries with most of the patient records obtained from Germany and the UK and a smaller number from France. IMS Disease Analyzer also directly includes

Table 15.1 Overview of GPRD, THIN, and IMS Disease Analyzer

	GPRD	THIN	IMS Disease Analyzer
Year data collection initiated	1987	2002	1992
Number of patients included	12.93 million total patients with11.39 million "research standard" patients contributing 67.0 million person years 5.14 million are currently "active patients"[2]	3.6 million "active" patients who can be followed prospectively, 10.0 million total patients contributing over 57.1 million patient years[5]	UK: 4.2 million total patients with 17.9 million patient-years Germany: 29.9 million (incl. 17.2 million German Specialty patients) with 54.5 million patient-years France: 5.2 million patients with 6.0 million patient-years[3,4,6]
Number of physician practices included	629[2]	498[5]	UK: 218 Germany: 2357 (plus 2010 specialty practices) France: 2091[3,4,6]
Coding used	READ, Multilex	READ, Multilex	ICD10, READ, ATC Codes
Software used	Vision	Vision	Various
Regular quality checks performed	Permanent ongoing data checks on all practices: Data quality assurance processes are undertaken as part of data processing. Patients are flagged as "acceptable" for use in research by a process that identifies and excludes patients with non-contiguous follow-up or patients with poor data recording that raises suspicion as to the validity of that patient's record. Up to standard dates (described in text)	For the first 100 practices, preliminary audits of consultations and prescriptions as compared to national levels were performed. Ongoing, all data collected undergo consistency and integrity checks. Feedback is provided to practices regarding UK quality metrics performances, medical history recording, and comparison of prevalence of disease with national levels where available. Acceptable mortality recording (described in text).	All data checked to meet quality standards and for plausibility. Feedback reports are given to each physician monthly, showing the physician's prescription patterns and those of colleagues within the IMS panel and within their specialty group.

data from some specialist groups in Germany: cardiologists, diabetologists, dermatologists, gynecologists, otolaryngology, neurologists, psychiatrists, pediatricians, urologists, and some surgeons, including orthopedists.[3,16]

Description

Data collection and structure

Each year, practitioners record information on 3–5 million patients for each of these databases, accounting for nearly 5–7% of the population within the UK in each of the databases (and 5–7% within Germany in Disease Analyzer).[3,15,17–22] Practitioners use the electronic medical record to document information about their patients from encounters and drugs prescribed. Data are extracted electronically from the medical record by specific software designed for this purpose, examined for completeness and accuracy by the database administrators, and then uploaded into the database in an anonymous form. Data are collected initially and then updated over differing intervals, adding information on new patients entering the system and updating the longitudinal profiles of existing patients. For example, THIN data are collected by In Practice Systems (INPS) electronically approximately three times per week. Identified errors are corrected prior to the data becoming available to investigators. Consistency checks are additionally run on the THIN data server to ensure consecutive collections are sequential and complete. GPRD uses a similar approach. IMS utilizes the support of cooperating software companies to collect data which are anonymized and then sent to IMS. Data are checked annually by the German Medical Association. All three databases abstract this data specifically for research purposes. Table 15.2 contains a detailed list of data collected in each database.

All three databases are representative of their respective populations in terms of age, sex, most diseases, and prescriptions written, meaning that the prevalence of these will be similar to the general population. In general, most regions of the represented countries are included, though the density of patients within each region in the database may not represent the exact proportion of people living in the region.[1,17,18,34] However, the reported frequency of some diseases and characteristics are not representative of the population. For example, there is variability in the recording of musculoskeletal diseases, including rheumatoid arthritis and osteoarthritis.[35] Similarly, the spectrum of socioeconomic status found in the databases may not reflect the true distribution in the country.[13,34]

Researchers utilizing the databases have access to anonymized patient medical history including co-morbidities. Diagnoses and symptoms are entered using diagnostic codes, described below. Additional information can sometimes be elicited by reading through anonymized free text entries. Most free text entries are available to researchers, and additional records can be requested for an associated cost for anonymizing data, with the cost differing by company.[2,4,36] Some practitioners may still keep paper record files, which may include precomputerization records, hospital discharge paperwork, or consultant letters. Both GPRD and THIN have additional data services that will obtain these data, for a fee, from the GP. In THIN this service is called Additional Information Services (AIS).

Laboratory tests, blood pressure, height, and weight are available in all three databases to varying degrees. For example, in both GPRD and THIN laboratory data from recent years is nearly complete but some older lab tests may not be available electronically if they were received by the GP in hard copy.[2] In IMS, hemoglobin A1c for diabetic patients is nearly complete but many other laboratory values are not recorded.[4] Prescriptions issued by the general practitioner are well captured in these databases, though not all prescriptions are linked to a diagnosis code.

Hospitalizations, referrals, and the resulting consultation letters are recorded to varying degrees among the databases. In GPRD, complete hospitalization data (including hospital specialist consultations) in England are automatically linked to the patient's record. In THIN, hospitalization data often depend on the GP manually entering this information, and mainly includes discharge date and

Table 15.2 Selected variables available for epidemiologic research

	GPRD	THIN	IMS Disease Analyzer
Health-care professional demographics	Can determine if nurse or doctor entering data	Can determine if nurse or doctor entering data	Age, sex, and years in practice of physician
Types of physicians	General practitioners	General practitioners	Mainly general practitioners, but in Germany also specialists, e.g., cardiology, gastroenterology, dermatology
Practice and patient demographics	*Practice* Region, practice size, practice-level socioeconomic status (index multiple deprivation), up-to-standard date, date of last registration, status of practice *Patient* Year of birth, sex, ethnicity (currently about 25% recorded, but also available via census data), socioeconomic class and other census data to small area level,[2] hospital and disease registry data. Additionally, approximately 60% of patients have practice level Index of Multiple Deprivation and Townsend scores through linkage in addition to post code derived socioeconomic status (GPRD internal data).	*Practice* Computerization date, Vision date, patients per practice, region sometimes provided *Patient* Year of birth for adults, month and year for children. Patient-level socioeconomic status (Townsend deprivation scores), region, ethnicity (20%)	*Practice* Region, community size and patients per practice, number of doctors, number of employees, emphasis (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).
Social history	Smoking (83–93%),[23,24] obesity (61–79%),[23,24,30] alcohol (around 80%)[23,25]	Smoking (86–94%),[26,27,31,32] obesity (73–83%),[32] alcohol intake (75–85%)[32]	Obesity (~40% had BMI)[28] smoking and alcohol recording unknown
Referrals and results of investigations	Linkage to hospital data (England) shows the majority of records and provides greater detail. Most labs are electronic.	Available electronically where referral is made using Vision though some may be in paper files. Most labs are electronic.	HbA1c, blood sugar, cholesterol, LDL and HDL are available but others are variably available but can be requested. Test results can be requested from paper files.

Therapy	Drug name, route, dose, frequency, duration, Immunizations including batch. Cost of drug can be added.	Drug name, route, dose, frequency and duration. Cost of drugs is available in linked file.	Drug name, route, dose, frequency, duration, cost of therapy.
Health-care utilization	Visit frequency, hospitalizations, and consultant visits. Links to hospital episode statistics (HES) provides detailed ward level resource utilization in England only.	All GP visits recorded, hospitalizations are entered by GP, not directly from hospital. Sick leave recorded if GP issues a note.	Visit frequency, hospitalizations, sick leave included.
Identification of pregnancy and families	Pregnancy and pregnancy outcomes. Mother–baby link via a family/ household algorithm/ family identification number.	Pregnancy recorded. Families are identified by a number given to each household.	Pregnancy variable except gynecologist records. Family documentation incomplete
Identification of death and cause of death	Date of death and cause(s) available via GPRD data and from linkage to Office for National Statistics central mortality data (ONS).	Death date recorded. If the cause is not recorded, death certificates can be requested for a fee. Some cause of death information is also available as a linked file. Acceptable Mortality Reporting (AMR) is a quality indicator given for the year in which mortality records are deemed complete.	Seldom recorded.
Available additional data (e.g., consult records, labs, paper files)	Hospital discharge summaries, consultant letters. All free text available.	Hospital discharge summaries, consultant letters. Most free text available.	Available upon request.
Questionnaires	Prospective data collections possible from both healthcare professionals and from patients. Response rates from three recent studies were about 90%[29] (and GPRD internal data).	Prospective data collections possible from both health-care professionals and from patients. Response to paper questionnaires is about 90%[33] (and THIN internal data).	Available upon request.

discharge diagnostic code. Hospitalizations are not recorded in the Disease Analyzer unless the patient was referred to the hospital by the GP. Consultant letters and referrals are also not available for the Disease Analyzer database. Referrals are captured in both GPRD and THIN, although consultations may be obtained in the form of hard copy letters. Finally, components of social history such as occupation are not routinely recorded.[13]

All three services, though more structured in GPRD and THIN, allow for questionnaires directed toward practitioners or patients,[33,37–39] to provide additional information about variables that are not available in the database or to augment the data provided in the database. Fees are paid to the database for administration of the questionnaire and practitioners receive a fee for questionnaire completion.

In all three databases, most data are entered using codes rather than free-text entries.[15,33,40,41] READ codes are a comprehensive clinical language developed in the UK utilizing standard alphanumeric codes to record patient diagnoses, symptoms, laboratory and radiographic tests, and processes of care (e.g., referrals). In the UK, Multilex codes encode drugs prescribed by the GP. Multilex codes are managed by First Data Bank, a private-sector company. The Disease Analyzer uses READ codes and International Classification of Disease 10th edition (ICD-10) codes in the UK and in the other two countries, and Anatomical Therapeutic Chemical classification (ATC) codes rather than Multilex codes for medications in Germany and France.[3] Creating a code list for specific medical conditions, symptoms, covariates, exposures, or drugs of interest to an epidemiologic study is very important for extracting the appropriate data from any of these databases. Methods for deriving such code lists have been described.[40]

Data quality: accuracy and completeness

Data completeness varies among variables and databases. Pregnancy, family structure, mortality, and cause of death are variably recorded and can be difficult to ascertain or can require the use of complicated coding algorithms (particularly for family structure).[42–51] Risk factors such as smoking and obesity may have gaps; before 2004 these data were often not recorded, though after the introduction of the UK national initiative of Quality and Outcomes Framework (QOF), there has been a substantial increase in the completeness of recording[52–56] of these and other variables. For example, prior to 2004, smoking was recorded for around 75% of patients in GPRD, whereas in a 2007 study, it was found to be recorded in nearly 90%.[31] Additionally, software improvements and quality improvement initiatives over the past 5 to 10 years have increased overall data capture. While overall data recording has improved, some information may not be captured in these databases. For example, medications mainly given by specialists and over-the-counter medications may be missing. However, the long-term use of medications also available over-the-counter, such as aspirin and non-steroidal anti-inflammatory drugs may be recorded.[57] In patients over the age of 60, chronically used, non-prescription medications seem likely to be captured given that the NHS provides free access to these medications when prescribed by the GP.

Data quality checks are performed by all three databases on ascertainment of the data at regular intervals on three levels as described above: (i) practitioner recording, (ii) data abstraction, and (iii) maintenance of the database. In general, each record is examined for presence of birth date, registration date, sex, and continuity of data recording. If a particular provider or practice regularly provides data in which these elements are missing, they will receive feedback on their performance and may even be dropped from the database altogether. When data are uploaded or abstracted from the medical records, the database company performs additional quality checks to make sure the data have been correctly uploaded or extracted. Patients who have transferred out of the practice or who have died are censored at that time but not removed from the database; the date of entry into the practice and date the patient left the practice are available. Finally, subsequent updates to the databases are verified for accuracy.[1,2,7,34,36] All three databases undergo routine updating of the software

used to collect, check, transfer, and present data. While this chapter was being written, the UK National Health System was undergoing major changes because of new leadership. New initiatives to improve data recording and data quality are expected.

Several quality measures encourage physician participation and accurate data collection. Contributing GPs receive monetary compensation and training in the use of their software, and regular evaluation of their prescribing behavior and data recording. Specific types of compensation differ depending on the database. Feedback reports are given to recording practitioners with tips on improving performance and, in some cases, a summary of the practitioner's prescribing habits relative to similar practices or across the UK. Other quality measures include audits of newly added practices and comparison of acquired data to national databases (e.g., mortality, hospitalizations, cancer, and cardiovascular registries).[3,43,58] GPRD includes direct linkage to several disease registries and national mortality reporting. THIN employs an additional quality measure known as "Acceptable Mortality Reporting" or AMR, denoting the year in which mortality reporting was deemed complete for each practice.[34,37] GPRD assigns each practice an "up-to-standard" date, the year in which data recording for that particular practice met GPRD standards for both patient data and completeness of electronic data recording in specified areas.[17]

The UK NHS also has made changes in recent years that have affected data quality. For example, Pay for Performance measures instituted in 2004 increased GP reliance on the electronic medical record, leading to more complete data recording, especially for specific medical conditions.[53,59,60] Pay for performance was designed to increase performance using 146 quality indicators for ten chronic diseases: asthma, cancer, chronic obstructive pulmonary disease, coronary artery disease, diabetes, epilepsy, diabetes, hypertension, hypothyroidism, mental health, and stroke.[61] Upon entering the medical record to begin an encounter, yellow quality indicator boxes appear for completion if a patient has one of these diseases.[61] In 2004–2005 under the new pay for performance program, 99% of patients with diabetes had a reported hemoglobin A1c value in the past 15 months, as compared to 87% in 1998. It is unclear whether reporting has also improved for diseases outside of the ten specified above.[62] Other quality improvement strategies within the UK include national standards for treatment of diabetes (2003) and heart disease (1999), incentives for cervical cytology and immunizations (early 1990s), and widespread use of audit and feedback to the GPs by Primary Care Trusts (1990s).[62]

Access to the databases

Access to the latest versions of the databases can be purchased through the following administrators: MHRA (www.gprd.com), THIN (http://csdmruk.cegedim.com/), and Disease Analyzer (www.imshealth.com). Datasets can be obtained on CD-ROM or via online access depending on the version purchased. For GPRD, the highest method of access is the GPRD GOLD (GPRD OnLine Data), which provides fast query access to assemble cohorts and associated data for download to any statistical package. Alternatively, the GPRD research team will cut and provide data. Similarly, available for purchase is access to the whole THIN dataset under license, a subset of data from the dataset, or a preprocessed dataset with some data manipulation. Finally, for Disease Analyzer, researchers can buy the software and data with monthly updates, preprocessed datasets, or analyzed data through IMS.

Studies must be first reviewed by the home institution's institutional review board (IRB) and the ethics board of each database. Given researchers' inability to identify individual patients, such studies often meet the criteria for IRB exemption. However, ethics approval must still be sought through the databases. For GPRD, ethics approval for observational studies by the Independent Scientific Advisory Committee (ISAC) is undertaken by GPRD through a standardized application. Similarly, THIN studies undergo scientific review by THIN's Scientific Review Committee (SRC). However, if additional information will be collected, ethics approval is also obtained through the UK's National Health System Multi-Centre Research

Ethics Committee. Approval for IMS Disease Analyzer studies in the UK is similar to GPRD and THIN, but studies in Germany require only home institution IRB approval. Requirements for approval change over time and therefore the investigator should check with the data vendor about approval requirements prior to initiating a study. The companies additionally require the completion of a data use agreement prior to initiation of the study.

Strengths

Population-based data and sample size

Population-based studies draw subjects from the greater population to arrive at a sample that is reflective of the source of individuals from which the sample was derived.[63] All three medical record databases allow researchers to use population-based study designs, minimizing selection bias and improving the validity and generalizability of epidemiologic studies. In these databases, whole practices are enrolled rather than individual patients, although patients can opt out of having their information used. Very few patients opt out; although the number is not known, it is suspected to be less than 0.1%.

Population-based data sources are ideal for case–control studies in which cases (e.g., individuals with disease) are all or a representative sample of all cases in a precisely defined population and controls are sampled randomly from the source population from which the cases were derived.[64] Similarly, population-based data allow for the design of cohort studies given the prospective data capture with long follow-up periods. As the data are largely representative of the general population, results are generalizable to the broader population.

Information about practices in which the patients are seen allows researchers to measure individual practice effects on health outcomes. Furthermore, the large number of patients with longitudinal follow up (millions of patients and millions of person-years of follow-up) allows for sufficient statistical power to study many rare outcomes.

Validity of information

The validity of the information in these databases has been extensively studied, with the highest number of validation studies in GPRD followed by THIN, and much a smaller number in Disease Analyzer. This wealth of information provides a major advantage over other types of databases. Studies of agreement between recording in the electronic medical record and capture of data (e.g., prescription medications and specialist referrals) have been performed for some of the databases.[3,15,17,42,43,65] Numerous studies have validated a variety of outcomes and diseases. If not previously performed, validation of the desired exposure and outcome to be studied should be performed prior to or as part of the study to ensure that a particular diagnostic code reflects the patient's true state. If a diagnostic code is not validated, spurious results could be obtained and the validity of the study compromised[15,33] (see also Chapter 41). In GPRD alone, 212 publications have reported on verification of 183 diagnoses including a range of conditions,[17,66,67] including atrial fibrillation,[68] cancer,[69] cataract,[70] chronic obstructive pulmonary disease,[71] autism,[66] inflammatory bowel disease,[72] lymphoma,[73] myocardial infarction,[67] Paget's disease,[74] rheumatoid arthritis and juvenile idiopathic arthritis,[75] pregnancy outcomes,[44,45] pressure ulcers,[76] psoriasis,[13,77,78] psychosis,[79] suicide,[80] and venous thromboembolism.[81] Similarly, a number of reports of validation studies in THIN have been published, including quality of cancer reporting,[1,82] non-melanomatous skin cancer,[38] stroke,[15] peptic ulcer disease,[15,32] colon cancer, myocardial infarction,[15] date of death and mortality reporting,[37] hepatitis C virus infection,[33] and psoriasis.[78] A few specific outcomes including venous thromboembolism[83] and a general validation of pharmacoepidemiology and pharmacoeconomic studies has been performed for the IMS Disease Analyzer.[3] The general validation by Becher *et al.* descriptively analyzed the data available in the German database and compared the findings to national statistics. The authors explored types of physicians in the database, several diagnoses (e.g., diabetes and numerous malignancies), prescribing behaviors, and patient insurance plans. These are a

few examples of validation studies for each of the three databases, but many more exist.

Access to original medical records

Obtaining data from the medical record allows for a complete overview of the patient's history. Most, if not all, data about the patient are funneled through the GP and therefore accessible to the researcher. Laboratory and radiology data are mostly available and therapy data are complete except for medications administered in the hospital or by specialists (e.g., chemotherapy) and over-the-counter medications, which may be absent or variably recorded. Notably, via requests to the database administrators, anonymized copies of paper records, more detailed patient history, and consultation letters are obtainable, allowing researchers to verify information captured elsewhere or obtain additional information. However, this can be a costly process. Response rates for medical record requests have been greater than 80–90% in published studies in THIN and GPRD, with the majority of requests being met for GPRD within 3 months.[84–86] Studies of time to receipt of questionnaire results have not been performed in THIN nor IMS Disease Analyzer. An additional benefit in GPRD is the availability of original electronic medical records from hospitalizations, cancer registry data, and cardiovascular registry data, and direct link to these data for each patient in specific geographical regions.[13,17] Linkage data are currently available for approximately two million patients. Virtually all hospitalization data are thought to be captured. Diagnostic codes are recorded by unit clerks coding from the medical charts and all labs and testing procedures are uploaded electronically. While publications detailing validity of the hospitalization data are not yet available, these studies are in progress.

Weaknesses

Completeness of data

As the data are derived from the patient's GP medical record, the investigator is relying on the GP's complete and accurate recording of the patient's history and events. Any information received by the GP from consultants, hospitalizations, or test results in hard copy would need to be manually entered into the electronic medical record, and thus may not be fully captured. Radiology and laboratory reports received in hard copy may not be entered in all cases (and may be more likely to be entered if abnormal).

In general, much of specialist care is missing from these databases, as they are designed to capture general practitioner activity. Exceptions include IMS Disease Analyzer's capture of data from specialist offices in Germany, complete specialist information captured in the hospitalization data in GPRD, and the GP recording of consultation data in THIN and GPRD, although all consultations may not be recorded, particularly specialist information during hospitalization.[42,43] In one GPRD study, 10% of data on specialist consultations were missing.[42,43] These are very old estimates, however.

Because codes for each chronic disease may not be repeated at each visit, episodes of care involving acute events may be better recorded than chronic diseases.[13,65,87] Non-significant medical events or medical problems that are no longer clinically active may not be documented.

Limited information on patient-based socioeconomic status is included in GPRD and THIN, although these data are not available in Disease Analyzer. Occupation and employment are rarely if ever recorded in any of the databases. THIN provides a patient-level measure of socioeconomic status derived from the patients' postal code of residence (around 95% of patients have this information), whereas GPRD provides a practice-level measure of socioeconomic status for nearly all practices and has recently added a patient-level measure for 60% of patients.

Data on confounding variables such as smoking, alcohol use, body mass index, and height are not available for all patients (see Table 15.2 for data on percent recording). Data on medications given during hospitalization, those medications restricted to specialist care, and hospital discharge medications may be particularly problematic (though patients generally only receive up to 2 weeks of medications upon discharge). In addition, not all

medications are linked to a particular diagnosis. Around 50% are linked in Disease Analyzer and, while medications are not directly linked to a diagnosis in GPRD one can use diagnoses recorded during the visit when the medication was prescribed. Medication compliance is often not recorded so notation of a prescription does not necessarily mean the medication was taken. Also, prescription records only capture prescriptions written but do not indicate which prescriptions were filled. However, in THIN and GPRD, refills or repeat prescriptions are recorded, as the prescription is the payment document. New prescriptions are generated when the current refills have been used and a new record is created.

Finally, although these are longitudinal databases, patients may only be in the database for a few years if they transfer out of the practice or if the practice ceases to participate in the database. Thus, studies of incident exposure in which patients need to be followed for many years may suffer from loss to follow-up over time.

Complexity and costs of computer hardware and software needed

The size and complexity of these databases requires adequate computer hardware and software as well as experienced data managers. Knowledge of the applications used for data access and loading must be obtained, and stable telecommunications links must be available. In some cases, use of the database may require 100–200 gigabytes of disk storage, though this varies with the size of the database, the data fields available for review, and whether data are available as individual records or tables with summary counts for indicator variables of interest. As with other databases, investigators need to use caution interpreting dates in these databases. The date of any recording in the medical files may reflect either the date of data entry, the date on which the observation was made or in the case of new registrations in general practice, dates may reflect entry of data obtained from previous practitioners or from previous recording systems. Researchers using these databases for epidemiologic research need to be fully aware of these issues. Various computer programs and knowledge of their use may be required by researchers (e.g., Oracle Discover or SAS-based interface, database programs including Visual Dbase, Microsoft Access, and statistical packages Stata, SAS, and SPSS).[13]

Particular applications

Medical record databases have been an abundant source of scientific data for epidemiologic and pharmacoepidemiologic investigations and have been used for numerous applications including studies using the case–control, case crossover, self-controlled case series, and cohort designs, signal detection, drug safety, health economics including cost effectiveness, assessment of the natural history of a disease, time to event analyses, and incidence studies. Examples of these studies are provided here, but comprehensive lists of studies performed using each database can be found on the respective websites. Numerous peer-reviewed manuscripts have been published, and many abstracts have been presented at international conferences using these three databases (greater than 700 from GPRD alone, over 150 in Disease Analyzer, and more than 200 in THIN).

Representative incidence and prevalence studies[68,74,88–95] include shoulder complaints in UK primary care,[96] newly diagnosed heart failure in primary care,[97] bullous pemphigoid, and pemphigus vulgaris.[98] Other epidemiologic studies include the natural history of disease (e.g., irritable bowel syndrome[99]), the risk of a particular outcome occurring[90,100–108] (e.g., lymphoma among inflammatory bowel disease patients,[73] myocardial infarction in patients with psoriasis,[103] and complications of diabetes[109]), defining associated conditions[76,110,111] (e.g., obesity and liver disease[112]), and patterns of diseases or symptoms and the rate of referral[71] (e.g., chronic pelvic pain[113,114]).

Hundreds of pharmacoepidemiologic studies have been published. These include studies assessing risks[69,80,86,108,115–125] and outcomes[126–132] of medication (e.g., risk of myopathy and myalgias by statin class[120]), safety and tolerability of medications,[19,119,133–138] studies of medication exposure and pregnancy outcomes,[48,139] and reduction of

morbidity or mortality by medications[140] or interventions such as vaccinations,[141,142] aspirin,[127] and anti-inflammatory medications.[143,144] Other pharmacoepidemiologic studies focus on persistence of medication use in the primary care setting[145,146] (e.g., antihypertensives,[147,148] bisphosphonates,[149,150] and glaucoma therapies[151]), compliance with prescriptions,[152] physician's use of guidelines in prescribing medications[153] (e.g., antibiotics in children,[154] antidepressants[155]), and trends in prescribing.[156–165]

Finally, pharmacoeconomics and health services studies have used medical record databases for population-based studies.[166,167] Cost-effectiveness of bisphosphonates in elderly women,[168] comparison of cost between glaucoma therapies,[151,169] cost-effectiveness of long-term hormonal contraception,[170] and the cost-effectiveness of treatment of gastroesophageal reflux disease[171] are a few examples.[172–174] Health services researchers have examined health insurance-related barriers in obtaining new medicines and delay in access to new medical therapies,[166,175,176] health-care utilization in fibromyalgia[177] and diabetes,[178–181] prescribing trends and their budget impact,[182,183] equivalent care of the elderly and non-elderly in terms of investigating a symptoms concerning for ovarian cancer,[184] and vaccination uptake and distribution in the UK.[185] As one can see, the diversity of studies and variety of disciplines utilizing these databases are immense.

The future

It is important to recognize that changes in the national health systems may have an impact on data collection, quality, and variables included. For example, Pay for Performance increased the recording of data with regard to the diseases of interest, particularly diabetes. GPRD and THIN are working on data quality initiatives including practice level data quality indicators. GPRD has created linkage to death certificates, other health-care databases, and national registry data (e.g., hospitalization, cancer, cardiovascular, and mortality registries); although published data have not yet resulted, studies are underway. THIN has recently added pop-up questionnaires that appear for completion while the GP is charting within the medical record. When an investigator is requesting additional information, this feature allows the GP to complete the investigator's questionnaire at the point of care. GPRD has added the ability to perform genetic studies allowing for collection of blood samples, and interventional studies in which patients are randomized at the point of care. GPRD is also developing a new data collection system for new to the market drugs. This system already has permission to collect full prior and future electronic data on patients prescribed new to the market medications for over 15% of the UK population. IMS has developed several new information databases over the past few years including IMS Contract Monitor which contains information regarding the volume of drug delivery by public pharmacies, the health insurance plans, and the drug manufacturers. This information can then be merged with patient information from the Disease Analyzer. More developments like those mentioned here can be expected in the future. It is also important to note that studies utilizing these developments have not yet been published.

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. GPRD website. www.gprd.com (accessed August. 2011).
3. 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 Pharmacol Ther* 2009; **47**: 617–26.
4. IMS health website. www.imshealth.com (accessed Aug. 2010).
5. Cegedim. *Our data*. http://csdmruk.cegedim.com/our-data/our-data.html (accessed August 2011).
6. *IMS Disease Analyzer*. http://www.imshealth.de/sixcms/detail.php?id=1019 (accessed Aug. 2010).
7. deLusignan S, Stephens P, Naeema N, Majeed A. Does feedback improve the quality of computerized

medical records in primary care? *J American Medical Informatics Assoc* 2002; **9**: 395–403.
8. Royal College of General Practitioners. www.rcgp.org.uk (accessed Oct. 2010).
9. Heath D. France's health care system. *CMAJ* 2008; **178**: 596.
10. Lopes C. Health care in France: facing hard choices. *CMAJ* 2007; **177**: 1167–9.
11. 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**: 31–7.
12. Schlette S, Lisac M, Blum K. Integrated primary care in Germany: the road ahead. *Int J Integr Care* 2009; **9**: e14.
13. Gelfand J, Margolis D, Dattani H. The UK General Practice Research Database. In: Strom B, ed. *Pharmacoepidemiology*, 4th edn. John Wiley & Sons, 2005, pp. 337–46.
14. THIN Partnership. http://www.thin-uk.com/partnership.htm (accessed Aug. 2010).
15. Lewis J, Schinnar R, Bilker W, Wang X, Strom B. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2007; **16**: 393–401.
16. Dietlein G, Schroder-Bernhardi D. Use of the Mediplus patient database in healthcare research. *Int J Clin Pharmacol Ther* 2002; **40**: 130–3.
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**: 4–14.
18. Wong I, Murray M. The potential of UK clinical databases in enhancing paediatric medication research. *Br J Clin Pharmacol* 2005; **59**: 750–5.
19. Wood L, Martinez C. The General Practice Research Database: role in pharmacovigilance. *Drug Safety* 2004; **27**: 871–81.
20. Wood L, Coulson R. Revitalizing the General Practice Research Database: plans, challenges, and opportunities. *Pharmacoepidemiol Drug Saf* 2001; **10**: 379–83.
21. Lawson D, Sherman V, Hollowell J. The General Practice Research Database. Scientific and Ethical Advisory Group. *Q J Med* 1998; **91**: 445–52.
22. Walley T, Mantgani A. The UK General Practice Research Database. *Lancet* 1997; **350**: 1097–9.
23. GarciaRadriguez L, Tolosa L, Ruigomez A, Wallander M. Rheumatoid arthritis in UK primary care: incidence and prior morbidity. *Scand J Rheumatol* 2009; **38**: 173–7.
24. Mulnier H, Seaman H, Raleigh V, Soedamah-Muthu S, Colhoun H, Lawrenson R. Mortality in people with type 2 diabetes in the UK. *Diabet Med* 2006; **23**: 516–21.
25. 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**: 1–8.
26. Szatkowski L, Lewis S, McNeill A, Coleman T. Is smoking status routinely recorded when patients register with a new GP? *Fam Pract* 2010; **27**: 673–5.
27. Langley T, Szatkowski L, Gibson J, Huang Y, McNeill A, Coleman T, *et al*. Validation of the Health Improvement Network (THIN) primary care database for monitoring prescriptions for smoking cessation medications. *Pharmacoepidemiol Drug Saf* 2010; **19**: 586–90.
28. 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. *Hum Reproduct Update* 1999; **5**: 688–706.
29. Gonzalez-Perez A, GarciaRodriguez L. Prostate cancer risk among men with diabetes mellitus. *Cancer Causes Control* 2005; **16**: 1055–8.
30. 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.
31. 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.
32. Cai S, GarciaRadriguez L, Masso-Gonzalez E, Hernandez-Diaz S. Uncomplicated peptic ulcer in the UK: trends from 1997–2005. *Aliment Pharmacol Ther* 2009; **30**: 1039–48.
33. LoRe 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.
34. 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.
35. Jordan K, Clarke AM, Symmons DP, Fleming D, Porcheret M, Kadam UT, *et al*. Measuring disease

prevalence: a comparison of musculoskeletal disease using four general practice consultation databases. *Br J Gen Pract* 2007; **57**: 7–14.
36. THIN website. www.thin-uk.com (accessed Aug. 2010).
37. Hall G. Validation of death and suicide recording on the THIN UK primary care database. *Pharmacoepidemiol Drug Saf* 2009; **18**: 120–31.
38. Meal A, Leonardi-Bee J, Smith C, Hubbard R, Bath-Hextall F. Validation of THIN data for non-melanoma skin cancer. *Quality in Primary Care* 2008; **16**: 49–52.
39. Khan N, Harrison S, Rose P. Validity of diagnostic coding within the General Practice Research Database: a systemic review. *Br J General Practice* 2010; **60**: e128–36.
40. Dave S, Peterson I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf* 2009; **18**: 704–7.
41. 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**: 1.
42. Jick H, Jick S, Derby L. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ* 1991; **302**: 766–8.
43. Jick S, Kaye J, Vasilakis-Scaramozza C, Garcia Rodriguez LA, Ruigomez A, Meier CR, *et al*. Validity of the General Practice Research Database. *Pharmacotherapy* 2003; **23**: 686–9.
44. Charlton R, Cunnington M, Vries Cd, 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 Safety* 2008; **31**: 39–51.
45. 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**: 45–50.
46. Farmer R, Todd J. Pregnancy following oral contraception. *13th Intl Conf Pharmacoepidemiology, Orlando, August 1997*. 1997.
47. 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 Companion J Clin Psychiatry* 2010; 12.
48. Tata L, Lewis S, McKeever T, Smith CJP, Doyle P, Smeeth L, *et al*. Congenital malformation in children born to women with asthma. *Pharmacoepidemiol Drug Saf*. 2007; **16**(S271).
49. 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.
50. 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**: 555–64.
51. 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**: 758–62.
52. Health Act 2009. http://www.dh.gov.uk/en/Publicationsandstatistics/Legislation/Actsandbills/DH_093280 (accessed June 2011).
53. Department of Health. *Quality and Outcomes Framework.* http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4088693.pdf (accessed Feb. 2010).
54. Lester H. The UK quality and outcomes framework. *BMJ* 2008; 337: a2095.
55. 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**: 1–13.
56. Wise J. Framework for NHS annual "quality accounts" is launched. *BMJ* 2009; **339**: b3890.
57. Yang Y-X, 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.
58. Cegedim. *Quality assurance*. http://csdmruk.cegedim.com/our-data/data-quality.html (accessed June 2011).
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. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, *et al*. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006; **355**: 375–84.

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**: 181–90.
63. Szklo M. Population-based cohort studies. *Epidemiologic Reviews* 1998; **20**: 81–90.
64. Rothman K, Greenland S. *Modern Epidemiology*. Philadelphia: Lippincott-Raven, 1988, pp. 93–115.
65. Hollowell J. The General Practice Research Database: quality of morbidity data. *Popul Trends* 1997; **87**: 36–40.
66. 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.
67. Hammad T, McAdams M, 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**: 1197–201.
68. 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.
69. Jick H. Calcium-channel blockers and risk of cancer. *Lancet* 1997; **349**: 1699–700.
70. Derby L, Maier W. Risk of cataract among users of intranasal corticosteroids. *J Allergy Clin Immunol* 2000; **105**: 912–16.
71. 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.
72. 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.
73. 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.
74. 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**: 465–71.
75. 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**: 1314–21.
76. 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**: 321–5.
77. 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**: 1537–41.
78. Seminara NM, Abuabara K, Shin DB, Langan SM, Kimmel SE, Margolis D, *et al*. Validity of The Health Improvement Network (THIN) for the study of psoriasis. *Br J Dermatol* 2011; **164**: 602–9.
79. Nazareth I, King M, Haines A, Rangel L, Myers S. Accuracy of diagnosis of psychosis on general practice computer system. *BMJ* 1993; **307**: 32–4.
80. Jick S, Dean A, Jick H. Antidepressants and suicide. *BMJ* 1995; **310**: 215–18.
81. 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**: 591–6.
82. Arellano F, Conde E, Wenworth C, Schlienger R, Fernandez-Vidaurre C, Arana A. Validation of cases of lymphoma in THIN. *Pharmacoepidrmiol Drug Saf* 2008; **17**: S87–8.
83. 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 Reproduction* 1999; **14**: 1500–5.
84. 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**: 1111–16.
85. Bourke A, Dattani H, Robinson M. Feasibility study and methodology to create a quality evaluated database of primary care data. *Informatics in Primary Care* 2004; **12**: 171–7.
86. VanStaa 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**: 359–66.
87. 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.

88. Linsell L, Dawson J, Zondervan K, Randall T, Rose P, Carr A, *et al*. Prospective study of elderly people comparing treatments following first primary care consultation for a symptomatic hip or knee. *Family Practice* 2005; **22**: 118–25.
89. 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**: 2105–8.
90. 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**: 198–200.
91. Gupta D, Hansell A, Nichols T, Duong T, Ayres J, Strachan D. Epidemiology of pneumothorax in England. *Thorax* 2000; **55**: 666–71.
92. 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**: 1198–205.
93. 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**: 1209–16.
94. 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**: 304–10.
95. 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**: 612–18.
96. Linsell L, Dawson J, Zondervan K, Rose, P, Randall T, Fitzpatrick R, *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.
97. Johansson S, Wallander M, Ruigomez A, Garcia-Rodriguez L. Incidence of newly diagnosed heart failure in UK general practice. *Eur J Heart Fail* 2001; **3**: 225–31.
98. 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.
99. Ruigómez A, Wallander M, Johansson S, García-Rodríguez L. One-year follow-up of newly diagnosed irritable bowel syndrome patients. *Aliment Pharmacol Ther* 1999; **13**: 1097–102.
100. Kaplan G, Hubbard J, Korzenik J, Sands BE, Panaccione R, Ghosh S, *et al*. The inflammatory bowel diseases and ambient air pollution: a novel association. *Am J Gastroenterol* 2010; **105**: 2412–9.
101. 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**: 2784–8.
102. 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.
103. Gelfand J, Neimann A, Shin D, Wang X, Margolis D, Troxel A. Risk of myocardial infarction in patients with psoriasis. *JAMA* 2006; **296**: 1735–42.
104. Pujades-Rodriguez M, Smith C, Hubbard R. Inhaled corticosteroids and the risk of fracture in chronic obstructive pulmonary disease. *Q J Med* 2007; **100**: 509–17.
105. 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**: 1181–7.
106. Hernán M, Jick S, Logroscino G, Olek M, Ascherio A, Jick H. Cigarette smoking and the progression of multiple sclerosis. *Brain* 2005; **128**: 1461–5.
107. 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**: 1140–7.
108. VanStaa 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.
109. Huerta C, Zhao S, GarciaRodriguez L. Risk of acute liver injury in patients with diabetes. *Pharmacotherapy* 2002; **22**: 1091–6.
110. 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.
111. 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**: 723–7.
112. 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.
113. 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**: 1156–61.
114. 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**: 1149–55.
115. 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**: 314–17.
116. 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**: 758–62.
117. VanStaa T, Abenhaim L, Monette J. Rates of hypoglycemia in users of sulfonylureas. *J Clin Epidemiol* 1997; **50**: 735–41.
118. 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 Therap* 2004; **42**: 260–6.
119. 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.
120. 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**: e2522.
121. Zhang H, Carcia-Rodriguez L, Hernandez-Diaz S. Antibiotic use and the risk of lung cancer. *Cancer Epidemiol* 2008; **17**: 1308–15.
122. 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 Safety* 2008; **31**: 399–407.
123. 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.
124. Meier C, Derby L, Jick S, Vasilakis C, Jick H. Antibiotics and risk of subsequent first-time acute myocardial infarction. *JAMA* 1999; **281**: 427–31.
125. Meier C, Wilcock K, Jick S. The risk of severe depression, psychosis or panic attacks with prophylactic antimalarials. *Drug Safety* 2004; **27**: 203–13.
126. Edwards C, Cooper C, Fisher D, Field M, VanStaa 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**: 1151–7.
127. 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.
128. 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 Practice* 2008; **8**: 253–62.
129. 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. *Exp Clin Endocrinol Diab* 2006; **114**: 520–6.
130. 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**: 99–109.
131. Thomas M, Williams A. Are outcomes the same with all dry powder inhalers? *Int J Clin Pract* 2005; **59** (Suppl 149): 33–5.
132. 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.
133. 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 Safety* 2009; **32**: 1089–96.
134. Hubbard R, Lewis S, West J, Smith C, Godfrey C, Smeeth L, *et al*. Bupropion and the risk of sudden death: a self-controlled case-series analysis using The Health Improvement Network. *Thorax* 2005; **60**: 848–50.
135. 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 Inf Med* 2005; **44**: 697–703.

136. 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.
137. 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.
138. Tata L, Fortune P, Hubbard R, Smeeth L, Hawkey CJ, Smith CJ, *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.
139. 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.
140. Gonzalez M, Garcia-Rodriguez L. Proton pump inhibitors reduce the long-term risk of recurrent upper gastrointestinal bleeding: an observational study. *Aliment Pharmacol Ther* 2008; **25**: 629–37.
141. Schembri S Morant S, Winter JH, MacDonald T. Influenza and pneumococcal vaccination protect against all cause mortality in patients with COPD. *Thorx*. 2009; **64**: 567–72.
142. Bremner S, Carey I, DeWilde S, Richards N, Maier W, Hilton S. Vaccinations, infections and antibacterials in the first grass pollen season of life and risk of later hayfever. *Clin Exp Allergy* 2007; **37**: 512–17.
143. Belsey J. Non-steroidal anti-inflammatory induced upper gastrointestinal event rates in patients awaiting joint replacement in the United Kingdom. An epidemiologically-based burden of disease model. *Curr Med Res Opin* 2003; **19**: 306–12.
144. Rodríguez LG, Varas C, Patrono C. Differential effects of aspirin and non-aspirin nonsteroidal anti-inflammatory drugs in the primary prevention of myocardial infarction in postmenopausal women. *Epidemiology* 2000; **11**: 382–7.
145. Arellano F, Wentworth C, 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** (Suppl 2): S1.
146. 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**: 1311–8.
147. 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.
148. 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.
149. Cramer J, Gold D, Silverman S, Lewiecki E. A systematic review of persistence and compliance with bisphosphonates for osteoporosis. *Osteoporos Int* 2007; **18**: 1023–31.
150. VanStaa T, Geusens P, Leufkens H, Cooper C. Persistence to bisphosphonate treatment in actual clinical practice. *Bone* 2005; **36** (S2): S412.
151. 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**: 1519–27.
152. Schroder-Bernhardi D, Dietlein G. Compliance with prescription recommendations by physicians in practices. *Int J Clin Pharm Therap* Nov 2001; **39**: 477–9.
153. 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**: 1038–42.
154. Thompson P, Spyridis N, Sharland M, Gilbert RE, Saxena S, Long PF, *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.
155. 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 Safety* 2005; **28**: 1151–7.
156. 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**: 756–62.
157. Arellano F, Yood M, Wentworth C, Oliveria SA, Rivero E, Verma A, *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.

158. Boyc K, Yurgin N. Trends in the prescription of antidiabetic medications in France: evidence from primary care physicians. *Adv Ther* 2007; **24**: 803–14.
159. 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 Therap* 2006; **28**: 1144–54.
160. Dietlein G, Schroder-Bernhardi D. Doctors' prescription behaviour regarding dosage recommendations for preparations of kava extracts. *Pharmacoepidemiol Drug Saf* 2003; **12**: 417–21.
161. Schroder-Bernhardi D, Dietlein G. Lipid-lowering therapy: do hospitals influence the prescribing behavior of general practitioners? *Int J Clin Pharm Therap* 2002; **40**: 317–21.
162. MacDonald T, Morant S. Prevalence and treatment of isolated and concurrent hypertension and hypercholesterolaemia in the United Kingdom. *Br J Clin Pharmacol* 2008; **65**: 775–86.
163. Turner S, Thomas M, Ziegenweidt JV, Price D. Prescribing trends in asthma: a longitudinal observational study. *Arch Dis Child* 2009; **94**: 16–22.
164. Edwards C, Arden N, Fisher D, Saperia J, Reading I, VanStaa 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**: 1394–8.
165. 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.
166. 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**: 202–7.
167. Belsey J, Pittard J, Rao S, Urdahl H, Jameson K, Dixon T. Self blood glucose monitoring in type 2 diabetes: a financial impact analysis based on UK primary care. *Int J Clin Pract* 2009; **63**: 439–48.
168. VanStaa T, Geusens P, Leufkens H, Cooper C. An individualised model of cost-effectiveness of bisphosphonates in elderly women. *Bone* 2005; **36** (S2): S138.
169. 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**: 3009–16.
170. 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**: 1141–51.
171. Plumb J, Edwards S. Cost effectiveness analysis of proton pump inhibitors compared to omeprazole in the healing of reflux oesophagitis. *J Med Econ* 2002; **5**: 25–38.
172. Guest J, Hart W, Cookson R. Cost analysis of palliative care for terminally ill cancer patients in the UK after switching from weak to strong opioids. Palliative Care Advisory Committee. *Pharmacoeconomics* 1998; **14**: 285–97.
173. Poole C, Lebmeier M, Ara R, Rafia R, Currie C. Estimation of health care costs as a function of disease severity in people with psoriatic arthritis in the UK. *Rheumatology* 2010; **49**: 1949–56.
174. Guest J, Clegga J, Helter M. Cost-effectiveness of macrogol 4000 compared to lactulose in the treatment of chronic functional constipation in the UK. *Curr Med Res Opin* 2008; **24**: 1841–52.
175. 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**: 756–61.
176. Krobot K, Miller W, Kaufman J, Christ D. The disparity in access to new medication by type of health insurance: lessons from Germany. *Medical Care* 2004; **42**: 487–91.
177. 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**: 2489–99.
178. 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.
179. Rathmann W, Haastert B, Giani G. Drug prescriptions and costs in the treatment of diabetic polyneuropathy. *Ditsch Med Wschr* 1999; **124**: 681–6.

180. 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.
181. Gulliford M, Latinovic R, Charlton J. Diabetes diagnosis, resource utilization, and health outcomes. *Am J Manag Care* 2008; **14**: 32–8.
182. 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* 2008; **12**: 302–8.
183. Hughes D, McGuire A. The direct costs to the NHS of discontinuing and switching prescriptions for hypertension. *J Human Hypertension* Aug 1998; **12**: 533–7.
184. Tate A, Nicholson A, Cassell J. Are GPs underinvestigating older patients presenting with symptoms of ovarian cancer? Observational study using General Practice Research Database. *Br J Cancer* 2010; **102**: 947–51.
185. 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* Dec 2005; **27**: 371–7.

CHAPTER 16
In-hospital Databases

Brian T. Fisher,[1,2] Peter K. Lindenauer,[3,4] and Chris Feudtner[1,5]

[1] Center for Pediatric Clinical Effectiveness, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

[2] Divisions of Infectious Diseases, Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, 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

Introduction

Non-experimental studies are an important approach to assess the comparative effectiveness and safety of medications used in hospitalized patients. Ideally, clinical and medication data for the conduct of such studies should be ample and accurate to provide sufficiently precise and unbiased estimates of treatment effects, and also detailed regarding the temporal sequence of exposures and events to facilitate testing of hypotheses about cause–effect relationships. Furthermore, if the data originated from diverse hospital settings, the haphazard variation across settings in clinical practice that is not associated with patient characteristics could be exploited to compare the outcomes of similar patients who did and did not receive a given treatment.

The past two decades have witnessed the development of databases containing far more detailed daily information regarding inpatient care than what has historically been contained in typical administrative discharge data. These augmented databases have changed the analytic landscape: the compilation of such data from multiple institutions into a single database affords pharmacoepidemiologists the opportunity to conduct observational studies of inpatients across numerous hospitals. In this chapter, we aim to illustrate the strengths, limitations, and future prospect of these augmented databases by describing two illustrative examples—the Pediatric Health Information System database (PHIS) and the Premier Perspective™ Database (PPD)—which have been used as the basis for a variety of published pharmacoepidemiologic studies. Although the focus of this chapter is PHIS and PPD, other databases such as the University Health System Consortium comparative database and the Health Facts database (Cerner Corporation, Kansas City, Missouri) are also available for conducting pharmacoepidemiologic studies. These will each be discussed briefly.

Description

PHIS database overview

The Pediatric Health Information System (PHIS) is a comparative pediatric administrative database containing clinical and financial data elements for over 18 million patient encounters from 43 not-for-profit, tertiary children's hospitals in the US. Member hospitals represent most of the major metropolitan areas across the US. All hospitals submit inpatient cases, and 36 of the 43 submit emergency department, ambulatory surgery, and/or observation patients. PHIS data are updated on a quarterly basis and made available to each hospital through

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

a web-based reporting tool. Currently, data are readily available dating back to January 1, 2003, with archive data available from as early as 1992.

PHIS data typically come from two primary data sources within the participating hospitals. From the hospital's medical record system come the patient identification, demographics, dates of service, physician identification, discharge disposition, payer information, and up to 40 International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes and up to 40 ICD-9-CM procedure codes. From the hospital's billing system come everything the hospital provided and billed for by day of service. This resource utilization data includes all pharmaceuticals, laboratory tests (not results), imaging procedures (not results), supplies, room/ nursing charges, and other ancillary services. To make this charge-level data comparable, PHIS utilizes the Clinical Transaction Codes (CTC) to map each hospital's charge codes to a common classification system. The CTC system is owned and managed by Thomson Reuters Healthcare, the data processing partner.

Within PHIS each patient encounter is assigned a unique number. Each patient is also assigned a unique number. This number, a de-identified medical record number, allows a patient to be tracked across multiple admissions at a 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. Research groups have been established in the past to perform chart abstractions for the same study across multiple centers.[1,2] The PHIS data are updated on a quarterly basis and made available to each hospital through a web-based reporting tool.

Child Health Corporation of America

PHIS was created and is managed by the Child Health Corporation of America (CHCA). CHCA (www.chca.com) is owned cooperatively by the 43 free-standing, non-competing children's hospitals within the US that contribute to the database. CHCA supports a number of programs designed to improve the quality of care for children provided at each of these institutions, while at the same time improving the financial performance of each hospital. The contributing hospitals account for 85% of all free-standing children's hospitals within the US that are registered with the National Association of Children's Hospitals and Related Institutions (2010 data from The National Association of Children's Hospitals and Related Institutions, Alexandria, VA).

PHIS data quality and accuracy

Oversight of PHIS data quality and accuracy is a joint effort between CHCA, Thomson Reuters Healthcare, and participating hospitals. Each hospital uses a uniform file layout created specifically for PHIS with each data element having a detailed definition. When the files are submitted to Thomson Reuters the data are processed through a series of data quality audits. These audits primarily check for valid entries (e.g., valid ICD-9-CM diagnosis codes) and reasonable patient information (e.g., birth weight). Reports are generated that will identify errors in the data that exceed threshold levels at a particular hospital and thus need to be corrected for actual data upload. Specifically, two threshold levels are enforced: the first is the number of encounters from an institution that contain identified errors must not exceed 5%; second, the percent of billed charges with unmapped CTC codes must not exceed 2%. Error rates above either of these threshold values require hospitals to resubmit their data after errors are corrected and error rates fall below the threshold values. Additionally, each hospital must sign off on their data submission in order for their data to be included into the next PHIS data load. Data loads are performed quarterly and data are de-identified at the time of each upload.

Known data quality issues are communicated to all PHIS data users. These data quality reports allow the data users to appropriately exclude data for data quality reasons. Many data quality issues in PHIS are addressed by the participating hospitals, which are encouraged to resubmit historical data to correct known data quality issues.

Premier database

Overview of Premier

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 of these hospital and health systems. It currently serves more than 2300 academic medical center or community-owned hospitals, as well as more than 66 000 other health-care sites. Premier member hospitals are located across the US and range in size and setting from small, rural to large, inner-city hospitals.

Premier Perspective™ Database (PPD)

PPD was created and continues to be managed by Premier, Inc. PPD compiles hospital administrative data from approximately one-sixth of all hospitalizations in the United States. PPD represents hospitals that admit both children and adults.

The PPD contains information on more than 130 million patient discharges. Since its inception in 2000, Premier has collected and managed data from more than 593 hospitals. Currently, 66% of hospitals update their data each month, with the remaining hospitals updating quarterly. Data elements within each patient hospitalization record include demographic data, admission and discharge dates, admission diagnosis, all discharge diagnoses and procedure codes, discharge disposition, and extensive information regarding medication exposure (based on pharmacy billing records), laboratory testing, as well as supplies and services provided. Member hospitals access Premier data via web-based tools.

For most data elements, fewer than 1% of patient records are missing information, and key data elements, such as patient demographic and diagnostic information, have levels of missing data less than 0.01%. Premier member hospitals, which routinely generate charge and administrative data using a variety of coding schemes, first map all data elements to a standard Premier nomenclature and then transfer these data, recoded according to the nomenclature standards, to Premier. Next, Premier initiates data quality checking, with any identified data quality problems resulting in the return of the data to the allied hospital for resolution of the problem.

PHIS and PPD data structure and elements

The PHIS and PPD databases employ different data structures to organize essentially the same data elements. In PHIS, data for each patient admission are separated into eight distinct tables (patient abstract, diagnoses, procedures, pharmacy, clinical resources such as specialty consultations, non-pharmacy supply, laboratory testing, and radiographic imaging); and data within and across each table are linked to a specific admission by a unique discharge identification number. In PPD, four major tables (patient abstract, billing, CPT codes, ICD-9-CM codes) contain all data elements, with a variety of ancillary tables that facilitate the labeling of data elements.

The data elements of PHIS and PPD are mostly the same. While variable names and the values and value labels for specific variables differ, common data elements include:

- *Demographic and hospitalization data:* Data include patient age, gender, admission and discharge dates, insurance providers, summary of hospital charges and classification based on the all patient defined diagnosis-related groups (APR-DRG).
- *Diagnoses:* Admission and discharge diagnoses made during a particular hospital admission are reported by ICD-9-CM codes. In PHIS, prior to 2010, up to 21 distinct ICD-9 discharge diagnosis codes were listed for each admission; from 2010 onward, the upper limit is 41 codes. PPD allows an unlimited number of discharge diagnosis codes. Both PHIS and PPD are preparing for the transition to the tenth revision for the classification of discharge diagnoses (ICD-10-CM) in 2013.
- *Pharmacy:* Pharmacy billing data are utilized to establish the generic name, formulation (e.g., parenteral or enteral) and dose for each medication ordered for a patient for each hospital day, throughout the entire hospitalization. These billing data are generated from the each participating hospital's master charges which are detailed lists of chargeable items and services.

• *Procedures:* Each procedure performed for a patient during an admission is documented by ICD-9-CM procedure codes in both PHIS and PPD. PPD also contains Current Procedural Terminology (CPT) codes. Similar to the pharmacy data, the timing of each procedure is linked to the hospital day on which the procedure was billed and thus the actual day of the procedure is known.
• *Clinical and supply:* Provision of clinical resources (e.g., sub-specialty consultations) and non-pharmacy patient supplies (e.g., serial compression devices to prevent venous thromboembolism) are also documented and linked to a specific hospital day for each admission.
• *Laboratory and radiologic imaging:* Every laboratory and radiologic study that was billed for during the hospitalization is documented and associated with the day of the hospitalization on which the test was ordered. Laboratory and imaging results are not available as a part of these databases.
• *Charges:* All drugs and test, services, and supplies in the databases have the corresponding charges for each item, along with either actual charges or cost-to-charge ratio parameters.

Table 16.1 provides examples of common data elements.

Data use agreements compliant with the Health Insurance Portability and Accountability Act (HIPAA) as well as (when applicable) institutional review board-approved study protocols govern the proper use of these data. Data from either database is largely freely available for query and data extraction for member hospitals, or for non-member institutions, can be purchased at a cost dictated by the scope of data requested.

Table 16.1 Examples of data elements contained in the PHIS and PDD data sets

Data type	Selected data elements
Patient abstract	Date of birth Race Gender Admission date Discharge date APR-DRG* classification
Diagnoses	Discharge diagnosis based on ICD-9-CM codes Order of discharge diagnoses
Pharmacy	Medications ordered Route of administration Day of administration Pharmacy charge
Procedures	Procedures performed based on ICD-9-CM codes Date of procedure
Supply	Supply ordered Day supply delivered Supply charge
Laboratory	Lab ordered Day lab delivered Lab charge *Does not include actual lab results*
Radiologic imaging	Imaging procedure ordered Utilization of contrast media Day imaging procedure was ordered *Does not include the actual results of imaging studies*
Clinical	Clinical service provided Day service was provided Charge associated with service

*APR-DRG, all patient refined diagnosis related groups.

Strengths

Sample size

There are a paucity of effectiveness and safety data specific to the administration of medications and medical interventions to inpatients.[3,4] A significant percentage of pediatric and adult admissions have been associated with either off-label or inappropriate medication administration.[5,6] Fortunately, in recent years there has been an increase in the number of FDA-approved medications but this has been associated with a concomitant increase in inpatient pharmaceutical expenditures.[7]

The availability of inpatient administrative 'plus' databases such as PHIS and PPD can provide pharmacoepidemiologists with an opportunity to define and create datasets representing a large number of hospitalized patients from a multitude of medical

institutions across the United States (see also Chapter 28). PHIS alone represents 85% of all the freestanding children's hospitals in the United States and PPD captures approximately 15% of all pediatric admissions across the nation; together these two datasets are estimated to represent up to one-fifth of all the pediatric admissions in the United States. PPD reflects approximately 5% of all adult admissions in the United States. 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.

Versatile data source

The extensive nature of these databases allows the researcher to be versatile in approaching various pharmacoepidemiologic questions. The databases can and have been used to evaluate the treatments of, and resource utilization for, various medical conditions that are both common and rare.[2,8,9] Because they are nationally representative, the data derived from PHIS and PPD also afford the ability to examine geographic variations in, and treatment of, a specific illness.[10] Data from either database can be organized in a manner that links admissions longitudinally for a specific person. Formatting the data in such a way allows for the investigation of the impact of medication exposures over time. Furthermore, because the pharmacy data and medical intervention data are documented for each hospital day, the timing of a particular intervention or medication exposure can be ascertained. This knowledge of the daily exposure can help to establish a temporal association between an exposure and outcome, thus strengthening the implication of such an association.

Data quality

Both PHIS and PPD have data quality oversight ensuring that the data uploaded from each institution meet predefined quality standards. In the event that data are missing, attempts are made to reconcile these missing elements. In the circumstance that missing data elements cannot be reconciled, the data users are informed of the location and extent of the missing data. Therefore, researchers utilizing these databases are assured about the quality of available data and are well informed about 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).

Efficient and inexpensive

These databases are readily accessed either via a virtual network (PHIS) or with assistance from the database administrators (PHIS and PPD). By creating simple 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 manipulation usually is necessary to establish a dataset suitable for analysis. For example, the data can be easily transformed into a time-dependent format for survival analysis. For researchers at PHIS contributing institutions, these data are completely free. Data from PPD can be purchased for a reasonable price. The cost of data is negotiable depending on the extent of the desired dataset.

Data longevity

Both of these administrative "plus" databases have been collecting data for over a decade. As hospitalizations for individual patients can be linked over time, these databases afford the opportunity to analyze illnesses that may require frequent readmissions (e.g., malignancy or autoimmune conditions). Furthermore, the continuous collection of data will allow researchers to trend the impact of certain illnesses from year to year (e.g., methicillin resistant *Staphylococcus aureus*).

Weaknesses

PHIS and PPD have two principal limitations that weaken the strength of inference of studies based on these data sources. First, the generalizability of study findings can be questioned since the degree

to which the hospitals that contribute to either dataset differ from non-contributing hospitals, in terms of clinical practice or patient case mix, is not well characterized.

Second, the validity of findings may be compromised because of some degree of misclassification regarding several aspects of patient classification: (i) disease status, since the accuracy of clinical diagnoses, encoded as ICD-9-CM codes at the time of hospital discharge, cannot be readily validated across all hospitals, and these diagnostic codes used for billing may not reflect the comprehensive set of clinical diagnoses that were made for a given patient; (ii) exposure status, since drug exposure data is based on billing for dispensed drugs, which most likely but not inevitably were administered to the patient; and (iii) outcome status, for the same reason mentioned above regarding clinical diagnoses made during the hospitalization, as well as the possibility of the outcome occurring after hospital discharge and the patient either being readmitted to another hospital or not hospitalized. Furthermore, exposure and outcome status may be subject to an ascertainment bias since hospitalized patients have different lengths of stay, which can vary across different types of hospitals, resulting in different durations of observation for exposures and for outcomes.

The research questions that these datasets can address are also 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 were available. Similarly, information regarding the precise day within a hospitalization on which a specific diagnosis was made or complication arose, or a reliable and validated measure of the severity of a patient's medical condition at the time of admission to the hospital, would enable more sophisticated analysis to draw conclusions based on stronger inferences (note that the existing severity of illness index is based on diagnoses and procedures that occurred throughout the hospitalization and thus reflects the total hospitalization course of illness). These limitations, however, can be partly mitigated by reviewing patients' medical records to supplement the database with additional information of interest.

Particular applications

Despite the previously discussed limitations, the comprehensiveness of these multicenter databases make them attractive options for testing clinically important pharmacoepidemiologic hypotheses in hospitalized adult and pediatric patients, hypotheses that were previously limited by insufficient data. Pharmacoepidemiologists have already utilized these databases to publish important findings specific to the diagnosis and treatment of frequent and rare illnesses, variations in therapeutic practices, drug toxicity monitoring, and comparative effectiveness studies. Examples of both pediatric and adult studies for each of these categories are discussed below. Certainly, additional areas of pharmacoepidemiologic research, ranging from analysis of temporal trends and practice variation to monitoring of drug toxicity and comparative effectiveness evaluations of treatments, can and are explored in these databases as well.

Temporal trends of diagnosis and treatment

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.[11–13] Given the 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.[14–16] Furthermore, these trials had raised concern about hemorrhagic complications from the APC therapy. Additional randomized controlled trials are being performed. However, it will be several years until the data from these trials are available, leaving clinicians without further guidance on their current patients.

Therefore, a retrospective study using the PDD 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.[17] 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 2 days of admission.

The investigators identified 33 749 patients meeting the study eligibility criteria of which 4.7% received APC in the first 2 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 revealing a benefit of APC on mortality (OR of 0.87 95% CI: 0.80–0.95). Interestingly, the rates of gastrointestinal bleeding, intracranial hemorrhage, and major transfusions were similar between those treated with APC as compared to those not treated. While it has not negated the importance of currently active randomized controlled trials, this well-executed, non-randomized study has provided both timely and relevant data regarding the treatment of sepsis in routine clinical practice.

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. For the past two decades, the epidemiology of VTE has been hypothesized to be 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, predispose 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 3 and 2 years of data from the 1990s. Thus, both studies lacked any ability to detect significant temporal trends. 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 has 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.[18] 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 16.1a). 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 16.1b). These findings have focused 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.[19] Currently, no consensus therapeutic guideline addresses HSP outpatient or inpatient management, which may include treatment with corticosteroids, antihypertensives, and non-steroidal anti-inflammatory drugs (NSAIDs) or opioids, and the performance of various laboratory and radiographic tests.

A study of 36 children's hospitals from 2000 to 2007 with hospitalization records contained in PHIS sought to assess the variation in inpatient therapy of children with HSP.[9] 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,

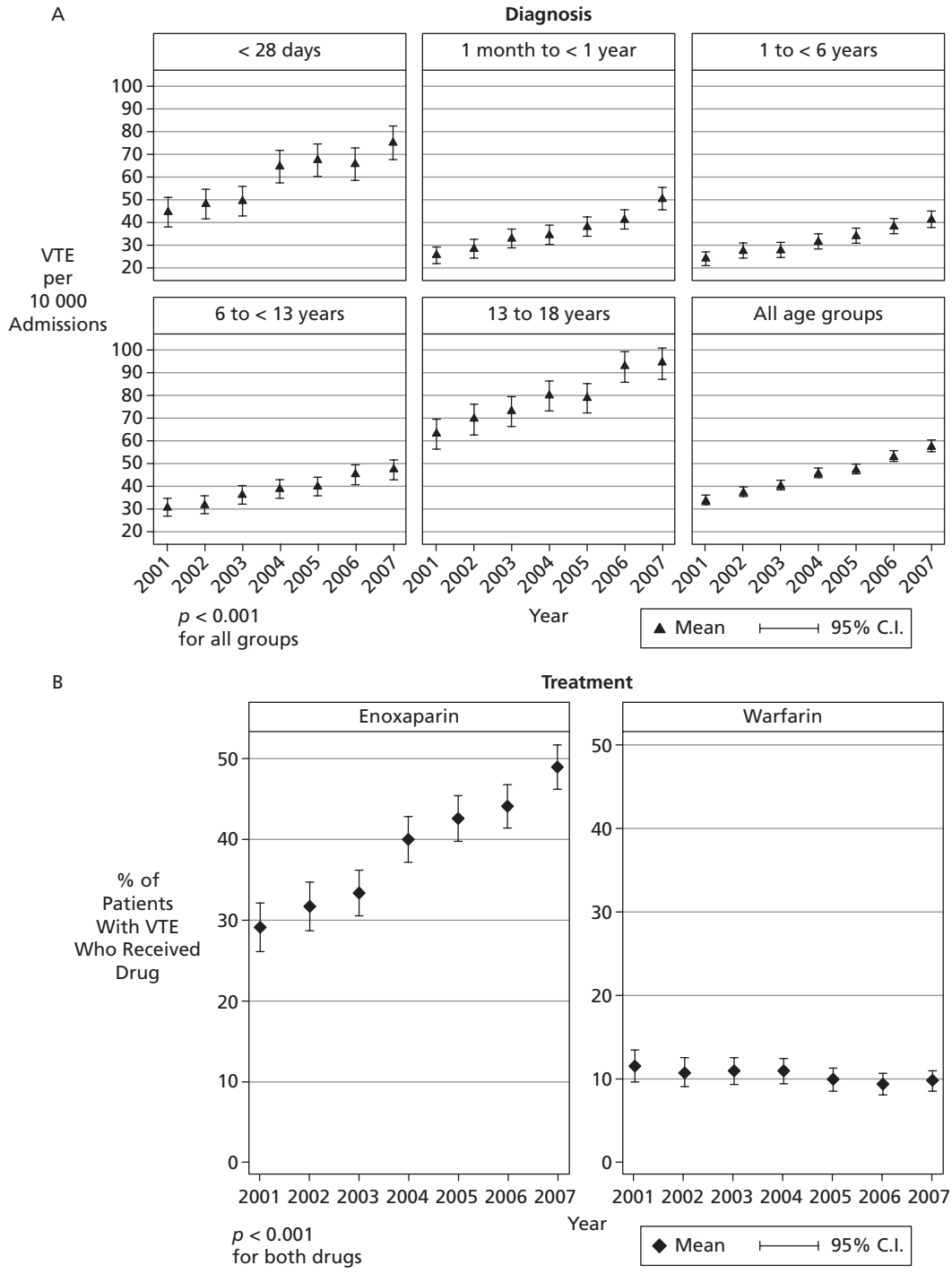


Figure 16.1 Temporal trends in diagnosis and treatment of venous thromboembolism (VTE). Adapted from Raffini *et al.*[18] with permission from the American Academy of Pediatrics.

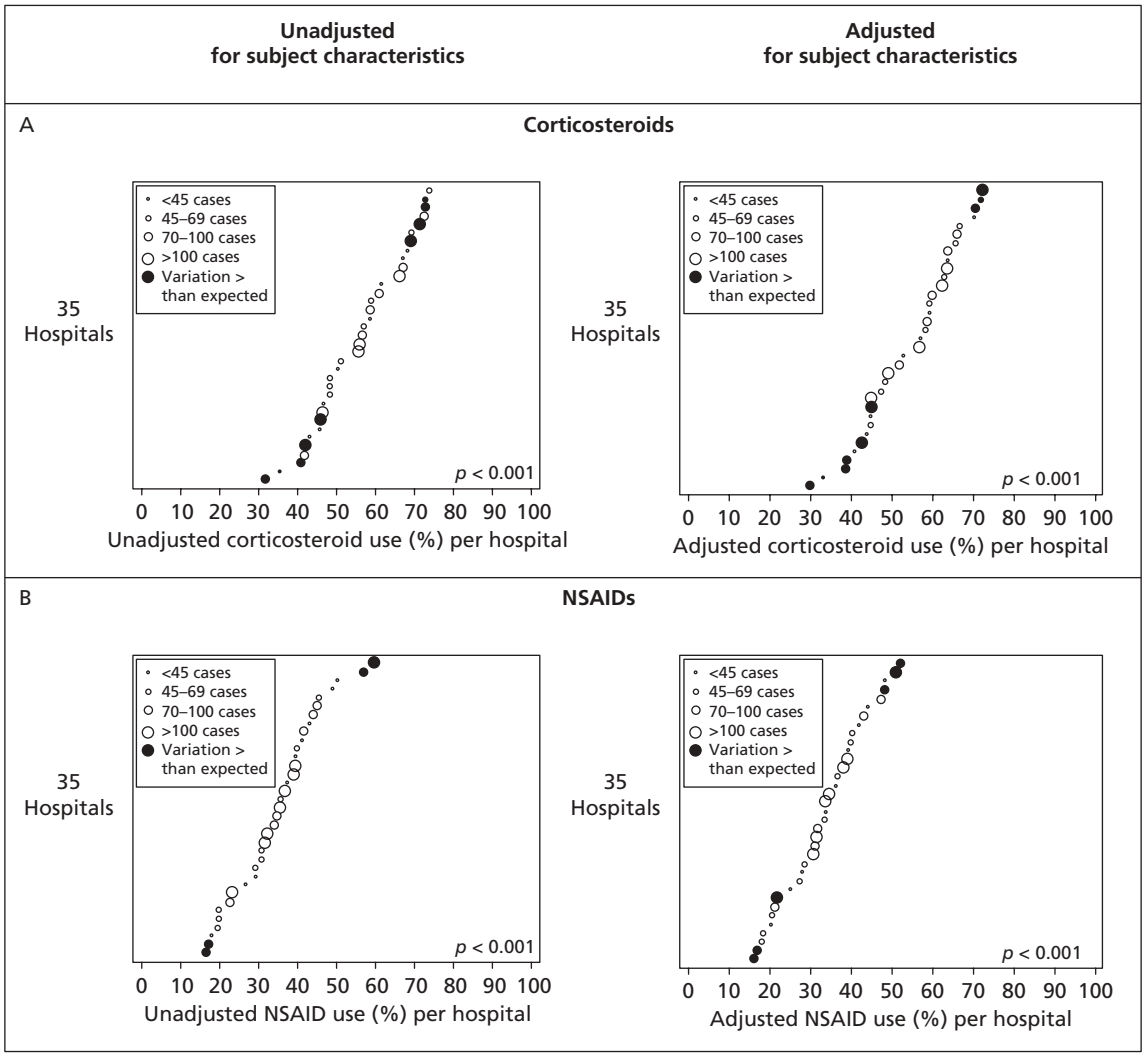


Figure 16.2 Variation in medication use among hospitals in unadjusted and adjusted models. Adapted from Weiss *et al.*[9] with permission from Elsevier.

35% NSAIDs, and 11% antihypertensive drugs. Substantial variation in the use of these medications was evident across the hospitals (Figure 16.2), and persisted despite adjustment for case-mix differences among hospitals. These findings have underscored the need for additional pharmacoepidemiologic outcomes research to enable the formulation of evidence-based practice guidelines.

Adverse event evaluation

Medication use in hospitalized elderly patients

Elderly patients respond to certain medications differently than younger patients, with either reduced drug efficacy or heightened susceptibility to adverse effects. Consequently, particular drugs are deemed best to be avoided when treating geriatric patients.

The Beers list, first developed in 1991 and subsequently revised, identifies such drugs to be avoided.[20] 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.

Rothberg *et al.*[21] used the PPD 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. 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 common PIMs were promethazine, diphenhydramine, propoxyphene, clonidine, amiodarone, and higher doses of lorazepam. Compared to internists, geriatricians were less likely to prescribe high-severity PIMs (AOR 0.69; 95% 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 one avenue 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.

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 the hospital for initiation of broad-spectrum antibiotics.[22] In many instances cefepime is the primary antibiotic utilized for these fever and neutropenia episodes. In 2007, a meta-analysis 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 as compared to those receiving other beta-lactam antibiotics.[23]

These data raised significant concern about the safety of cefepime. However, uncertainty existed regarding the true implications of the results from the meta-analysis 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 meta-analysis; lastly, a plausible theory for the increased risk for mortality secondary to cefepime was not identified. It was clear that additional data and analysis 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 16.2 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 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 PHIS.[24]

Table 16.2 Comparison of demographic variables from a pediatric AML cohort created retrospectively in PHIS and another cohort assembled in a prospective chemotherapy trial[24,25]

	Retrospective PHIS AML cohort (N = 917)	Prospective AML chemotherapy trial (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 years	186 (20.3%)	107 (22%)
2 to less than 16 years	603 (65.8%)	318 (65%)
Older than 16 years	128 (14.0%)	67 (14%)
Race		
White	649 (71%)	316 (64%)
Black	119 (14%)	42 (9%)

After extracting all admission information for each patient for up to 1 year from diagnosis, a survival dataset was created. After adjusting in a Cox regression 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 between patients recently exposed to cefepime versus those exposed to ceftazidime, an antipseudomonal penicillin, or carbapenem.[25]

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 meta-analysis. The FDA's own review and meta-analysis did not identify an increased risk of death associated with cefepime.[26] The above study highlights the opportunities that a pharmacoepidemiologist can capitalize on when using an administrative database such as PHIS. A relatively large cohort of children from across the United States with a rare illness was established in an efficient time frame. Within this cohort the safety of a medication was then considered in a time-dependent manner. These results had an immediate impact by providing pediatric clinicians reassurance in the continued use of cefepime in their patient population.

Comparative effectiveness

Antibiotics for chronic obstructive pulmonary disease

It has been estimated that as many as 24 million United States residents suffer from chronic obstructive pulmonary disease (COPD).[27] Acute exacerbations of COPD require inpatient care, resulting in as many 600 000 hospital admissions annually totaling an estimated $20 billion in direct costs each year.[28] Infection is a primary contributor for such exacerbations. Current COPD treatment guidelines suggest that antibiotics be initiated 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.[29–31] This recommendation for antibiotics is based on limited data from relatively small randomized trials, most of which were performed close to two decades ago.[32]

A retrospective cohort of patients hospitalized between January 1st, 2006 and December 31st, 2007 for an acute exacerbation of COPD was assembled using the PDD. Using this cohort the investigators compared the effectiveness of antibiotics in the first 2 days of admission to no antibiotics in those first 2 days.[27] 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). 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 47) to balance 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 treatment failure.

The cohort consisted of 84 621 patients, 79% of whom received antibiotics in the first 2 hospital days. In the propensity and covariate adjusted model the resultant odds ratio was 0.87 (95% CI: 0.82–0.92) favoring early antibiotic administration. This positive impact of early antibiotic use was most pronounced among those patients deemed to have the highest risk for treatment failure. Importantly, the presence or lack of increased sputum production or increase in dyspnea, as reflected in the use of sputum testing, did not alter the point estimate. This study identifies 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 current guidelines, antibiotics may be beneficial in all patients requiring hospitalization.

Although this study cannot be considered equivalent to that of 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 methodological 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.

Treatment of osteomyelitis

Osteomyelitis is a bacterial infection of the bone affecting one in 5000 children under the age of 13 and accounting for as much as 1% of all pediatric hospitalizations.[33,34] When not adequately treated, these infections of the bone can result in significant disability of the affected limb. Traditionally, prolonged courses of intravenous antibiotics (4–6 weeks) have been given as the standard of care for such infections. These prolonged courses of parenteral antibiotics necessitate venous access with some form of central venous catheter (CVC). Unfortunately, the use of a CVC presents a host of additional risks for the patient such as secondary infections, thrombosis, and line malfunction. These and other complications of CVCs were commonly observed in children receiving therapy for osteomyelitis.[35]

Over 30 years ago, pediatric studies suggested that well-absorbed, orally administered antibiotics could be utilized as a reasonable and less invasive alternative to prolonged intravenous therapy.[36,37] Based on these data some clinicians thought it reasonable to treat osteomyelitis with oral antibiotics while others thought the aforementioned studies were limited in their size and lacked compelling evidence for similar effectiveness between enteral versus parenteral antibiotics. Therefore, many clinicians have continued to support prolonged intravenous therapy as the *de facto* standard of care.

In order to evaluate the comparative effectiveness of oral versus prolonged intravenous antibiotics for the treatment of osteomyelitis, the PHIS database was used to assemble a retrospective cohort of 1969 children admitted for the treatment of acute osteomyelitis at 29 different hospitals.[2] An elegantly performed statistical analysis included the utilization of a propensity score (see Chapter 47) in a similar manner as discussed above to balance variables that may have contributed to the choice for route of antibiotic administration for each patient. There was no identified difference in the frequency of treatment failure between the two treatment strategies (5% for those intravenously treated versus 4% for those transitioned to oral therapy). Additionally, those patients given prolonged intravenous antibiotic therapy were

significantly more likely to suffer a treatment-related complication.

Certainly, the retrospective study discussed above is not the equivalent of a prospective randomized trial. However, the time and cost to perform such a trial with the resultant size of this retrospective cohort make such a study nearly impossible. Therefore, it is likely that such a randomized trial would never have been performed. This alternative approach has provided the pediatric medical community the necessary data that reveals that oral antibiotics are safer than, and at least as clinically effective as, prolonged intravenous therapy for the treatment of acute osteomyelitis.

Other databases

Both PHIS and PPD leverage hospital administrative and charge data to establish rich datasets that have and will continue to be used for pharmacoepidemiologic studies. Other databases such as Cerner's Health Facts database exist and have been used for similar purposes. The Health Facts database is a proprietary one and differs from PHIS and PPD in that it has merged clinical data (including admission, pharmacy, and laboratory data) from the electronic medical record with hospital claims data. This database currently represents 100 medical institutions across the US. It has been used in execute pharmacoepidemiologic studies such as the comparison of the in-hospital mortality of patients treated with analogue bolus insulin versus human bolus insulin.[38] The University HealthSystem Consortium (UHC) comparative database represents 107 primarily adult academic medical centers and their affiliated hospitals. It combines administrative, clinical, and financial data on each inpatient admission. This database contains risk-adjusted data that facilitates the comparison of outcomes between institutions. To date the published literature using this database has focused on surgical outcomes such as the comparison of laparoscopic surgery versus open surgery relative to the incidence of venous thromboembolism.[39] Not all of the member academic institutions share their complete charge data so this may limit the ability of this database to support pharmacoepidemiologic studies.

The future

In their current state, both PHIS and PPD can be used to perform important pharmacoepidemiologic studies on hospitalized adult and pediatric patients. The future impact of these databases can, nevertheless, be enhanced by further work in several areas.

First, the primary mechanism for identifying patients in the PHIS database with a specific complaint or illness is by way of ICD-9-CM admission or discharge diagnosis codes. Depending on the illness of interest the validity of the codes may not be known or may be less ideal for identification of patients. Chart abstraction can assess the accuracy of ICD-9-CM discharge diagnostic codes against the diagnoses and potential medical complications noted in the medical record, and test the assumption that medication and procedure billing accurately designates whether or not patients in fact received the medication or underwent the procedure. Other techniques, such as evaluating whether patients actually received the medications often used to treat a particular illness, based on primary medical records reviews, may strengthen the ability to identify patients with that disease.

Second, the data in both PHIS and PDD can be used to address additional important pharmacoepidemiologic and health research questions if the data are combined with other data, such as the results of inpatient laboratory or radiographic test results, patient-specific data from emergency department or ambulatory clinic encounters, or data regarding hospital's processes of care. For example, a study of infection control practices in children's hospitals was able, by supplementing PHIS data with data regarding each hospital's level of use of alcohol-based hand hygiene gel (gathered by a survey), to demonstrate a reduced odds of nosocomial gastrointestinal infections (adjusted odds ratio: 0.64; 95% confidence interval: 0.49, 0.85) among children cared for in hospitals were the hand gel was present.[40]

Third, the data from PHIS and PPD can be combined (only two hospitals are contained in both databases) to provide sampling coverage of both children's hospitals (as represented in PHIS and to a much lesser extent in PPD) and general hospitals (as represented only in PPD), thereby presenting a more complete and accurate representation of the entire realm of pediatric hospital care.

Lastly, the same pharmacoepidemiologic questions can be applied to and analyzed using each of the different 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.

References

1. Shaklee J, Zerr DM, Elward A, Newland J, Leckerman, K, Asti L, *et al*. Improving surveillance for pediatric *Clostridium difficile* infection: derivation and validation of an accurate case-finding tool. *Pediatr Infect Dis J* 2010; **30**: e38–40.
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**: 636–42.
3. Luo X, Doherty J, Cappelleri JC, Frush K. Role of pharmacoepidemiology in evaluating prescription drug safety in pediatrics. *Curr Med Res Opin* 2007; **23**: 2607–15.
4. Rommers MK, Teepe-Twiss IM, Guchelaar HJ. Preventing adverse drug events in hospital practice: an overview. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1129–35.
5. Shah SS, Hall M, Goodman DM, Feuer P, Sharma V, Fargason Jr. C, *et al*. Off-label drug use in hospitalized children. *Arch Pediatr Adolesc Med* 2007; **161**: 282–90.
6. Onder G, Landi F, Cesari M, Gambassi G, Carbonin P, Bernabei R. Inappropriate medication use among hospitalized older adults in Italy: results from the Italian Group of Pharmacoepidemiology in the Elderly. *Eur J Clin Pharmacol* 2003; **59**: 157–62.
7. Hoffman JM, Shah ND, Vermeulen LC, Schumock GT, Grim P, Hunkler RJ, *et al*. Projecting future drug expenditures—2006. *Am J Health Syst Pharm* 2006; **63**: 123–38.
8. 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**: 978–85.
9. Weiss PF, Klink AJ, Hexem K, Burnham JM, Leonard MB, Keren R, *et al*. Variation in inpatient therapy and diagnostic evaluation of children with Henoch Schonlein purpura. *J Pediatr* 2009; **155**: 812–18.
10. 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**: 243–7.
11. Karlsson S, Varpula M, Ruokonen E, Pettilä V, Parviainen I, Ala-Kokko TI, *et al*. Incidence, treatment, and outcome of severe sepsis in ICU-treated adults in Finland: the Finnsepsis study. *Intensive Care Med* 2007; **33**: 435–43.
12. Linde-Zwirble WT, Angus DC. Severe sepsis epidemiology: sampling, selection, and society. *Crit Care* 2004; **8**: 222–6.
13. 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**: 1546–54.
14. Abraham E, Laterre PF, Garg R, Levy H, Talwar D, Trzaskoma BL, *et al*. Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death. *N Engl J Med* 2005; **353**: 1332–41.
15. Bernard GR, Margolis BD, Shanies HM, Ely EW, Wheeler AP, Levy H, *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**: 2206–16.
16. Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, *et al*. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* 2001; **344**: 699–709.
17. 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**: 1101–7.
18. 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**: 1001–8.
19. Yang YH, Hung CF, Hsu CR, Wang LC, Chuang YH, Lin YT, *et al*. A nationwide survey on epidemiological

characteristics of childhood Henoch-Schonlein purpura in Taiwan. *Rheumatology (Oxford)* 2005; **44**: 618–22.

20. 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.
21. Rothberg MB, Pekow PS, Liu F, Korc-Grodzicki B, Brennan MJ, Bellantonio S, *et al*. Potentially inappropriate medication use in hospitalized elders. *J Hosp Med* 2008; **3**: 91–102.
22. Hughes WT, Armstrong D, Bodey GP, Bow EJ, Brown AE, Calandra T, *et al*. 2002 guidelines for the use of antimicrobial agents in neutropenic patients with cancer. *Clin Infect Dis* 2002; **34**: 730–51.
23. 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**: 338–48.
24. 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**: 3532–9.
25. 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**: 971–5.
26. Kim PW, Wu YT, Cooper C, Rochester G, Valappil T, Wang Y, *et al*. Meta-analysis of a possible signal of increased mortality associated with cefepime use. *Clin Infect Dis* 2010; **51**: 381–9.
27. 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**: 2035–42.
28. Snow V, Lascher S, Mottur-Pilson C. Evidence base for management of acute exacerbations of chronic obstructive pulmonary disease. *Ann Intern Med* 2001; **134**: 595–9.
29. National Collaborating Centre for Chronic Conditions. 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.
30. 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**: 932–46.
31. O'Donnell DE, Hernandez P, Kaplan A, Aaron S, Bourbeau J, Marciniuk D, *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.
32. 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.
33. Sonnen GM, Henry NK. Pediatric bone and joint infections. Diagnosis and antimicrobial management. *Pediatr Clin North Am* 1996; **43**: 933–47.
34. Vazquez M. Osteomyelitis in children. *Curr Opin Pediatr* 2002; **14**: 112–15.
35. Ruebner R, Keren R, Coffin S, Chu J, Horn D, Zaoutis TE. Complications of central venous catheters used for the treatment of acute hematogenous osteomyelitis. *Pediatrics* 2006; **117**: 1210–15.
36. Kolyvas E, Ahronheim G, Marks MI, Gledhill R, Owen H, Rosenthall L. Oral antibiotic therapy of skeletal infections in children. *Pediatrics* 1980; **65**: 867–71.
37. Tetzlaff TR, McCracken GH, Jr., Nelson JD. Oral antibiotic therapy for skeletal infections of children. II. Therapy of osteomyelitis and suppurative arthritis. *J Pediatr* 1978; **92**: 485–90.
38. Ko KJ, Tomor V, Nathanson BH, Bouchard JR, Aagren M, Dubois RW. Does type of bolus insulin matter in the hospital? Retrospective cohort analysis of outcomes between patients receiving analogue versus human insulin. *Clin Ther* 2010; **32**: 1954–66.
39. Nguyen NT, Hinojosa MW, Fayad C, Varela E, Konyalian V, Stamos MJ, *et al*. Laparoscopic surgery is associated with a lower incidence of venous thromboembolism compared with open surgery. *Ann Surg* 2007; **246**: 1021–7.
40. 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**: e387–92.

CHAPTER 17
Canadian Provincial Databases

Yola Moride[1] *and Colleen J. Metge*[2]

[1] Faculty of Pharmacy, Université de Montréal; Research Centre, University of Montreal Hospital Centre (CRCHUM), Montreal, Canada

[2] Faculty of Pharmacy, University of Manitoba, Winnipeg, Canada

Introduction

Canada, with its population of approximately 34 million, has a universal health-care program which, under the federal Canada Health Act of 1984, requires provinces to provide hospitalization and physician services without payment by the patient at the time of service. The administration of the program is under the responsibility of each of its ten provinces and three territories. From East to West the provinces, where the vast majority of residents are located, consist of: Newfoundland/Labrador (NFLD), Prince Edward Island (PEI), Nova Scotia (NS), New Brunswick (NB), Quebec (QC), Ontario (ON), Manitoba (MB), Saskatchewan (SK), Alberta (AB), and British Columbia (BC). Drug regulation is centrally conducted through Health Canada, an organization under the jurisdiction of the federal health ministry, but the administration of the drug coverage programs is conducted by the provinces and territories. Since public drug programs are not included in the Canada Health Act, the characteristics of drug coverage differ greatly across provinces. These in turn determine the nature of the study populations and the specific drugs that can be considered for pharmacoepidemiologic studies.

The health-services program includes physician visits, diagnostic tests, procedures, and hospitalizations, and covers all residents regardless of age or income. Physicians are paid on a fee-for-service basis, and databases have been created in each province for the administration of the program. A small number of physicians may have all or a portion of their activities covered by salary, and hence the services they provide may not be included in the medical services databases.

Prescription drug programs have been available for varying lengths of time in different provinces. Unlike coverage for physician visits, diagnostic tests, procedures, and hospitalizations, drug coverage differs across provinces, ranging from the entire population (e.g., universal in Saskatchewan and Manitoba) to specific segments of the population (e.g., elderly and welfare recipients in Ontario). In some provinces, the program has distinctive coverage features, described below.

Within each province, three health databases are available: (i) *medical services*, (ii) *hospitalizations*, and (iii) *prescription drugs*. The Saskatchewan databases have been reviewed in an earlier edition of this textbook.[1] These databases are linkable through a unique patient identifier that remains unchanged over time. In addition, it is possible to link this information to the demographic characteristics of the patients and to a variety of prescriber characteristics. Additional linkage capacities are available through province-specific databases, such as registries (e.g., cancer or cardiac registries), or through research initiatives, which are described later in this chapter. Through a study commissioned by the Office of Pharmaceuticals Management Strategies

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

(OPMS) of Health Canada (2010), an inventory of Canadian data sources relating to electronic information gathered in public repositories across Canada was conducted, and excerpts of the report are summarized in this chapter. The report, entitled *Data Sources to Support Research on Real World Drug Safety and Effectiveness in Canada: An Environmental Scan & Evaluation of Existing Data Elements*, was produced by Yola Moride and Colleen Metge for Health Canada and reproduction and distribution of the material is with the permission of Health Canada.[2]

The Canadian government is currently reforming its drug regulation by the implementation of progressive licensing. Under this framework, the current point-in-time licensing system will be replaced by a cyclical, progressive licensing model. This will be achieved through "the collection, analysis, and communication of knowledge and experience about a drug throughout its life cycle", including the postapproval setting.[3] This change in landscape has led to the creation of the Drug Safety and Effectiveness Network (DSEN) of the Canadian Institute for Health Research (CIHR), which is the Government of Canada's agency responsible for funding health research. The objectives of DSEN are "to increase the available evidence on drug safety and effectiveness available to regulators, policy-makers, health care providers and patients; and, to increase capacity within Canada to undertake high-quality post-market research in this area."[4] In this context, a large increase in the number of postapproval studies is forecasted.

This chapter provides a high-level review of the characteristics of the Canadian provincial databases, examples of linkage capacities and studies conducted through these databases, as well as strengths and weaknesses. Criteria that may be considered for the selection of a Canadian database are also put forward.

Description

Prescription drug databases

Eligibility criteria for prescription drug programs vary greatly across provinces. For example, in Ontario, the drug program is universal for all elderly residents over the age of 65 as well as for welfare recipients.[5] In Quebec the program also includes the vast majority of elderly (98%) and all welfare recipients.[6] In addition, since drug coverage is mandatory in Quebec since 1997, all residents and their dependents who do not have access to private insurance plans through their employers are covered by the public drug program. The public drug program imposes a deductible and co-payment, the amount of which depends on family income. In Alberta, the program also covers adults and children who are severely handicapped as well as those receiving palliative care.[7] In British Columbia, the public drug program was expanded in 2003 to cover patients with AIDS and cystic fibrosis as well as prescriptions dispensed in mental health services centers.[8] In Newfoundland/Labrador, patients with growth hormone deficiency are covered by the public drug program.[9] Nova Scotia also offers coverage for seniors, welfare, and patients with cancer.[10] In some provinces (e.g., British Columbia, Alberta, and Newfoundland/Labrador), expensive drugs may be covered by the public drug program but the level of co-payment depends on income, and may reach as much as 10% of a person's annual net income.[9] In Manitoba,[11] access to the public drug program depends on family income relative to the cost of drugs, and does not exceed 3.5% of a family's gross income. Saskatchewan[12] offers a drug plan to all residents, and co-payment also depends on annual income. Depending on the province, prescription data for subgroups of the population who are neither elderly nor welfare recipients may therefore appear either in public drug program databases or in private insurance databases. Such coverage features may affect the generalizability of findings obtained in studies conducted in adults between the ages of 18 and 65. The availability of longitudinal data may also be compromised by residents who migrate between the public and private insurance programs. However, dates of membership are available and continuous membership in the public drug program may be used as a study eligibility criterion. Two exceptions are British Columbia and Manitoba. In the former, a collaborative program has been implemented (BC PharmaNet) under the

auspices of the Ministry of Health, which combines data on all prescriptions dispensed through the public drug program, private insurers, and out-of-pocket.[13] Essentially, all prescriptions dispensed in community pharmacies regardless of coverage are centrally recorded in BC PharmaNet. In Manitoba, prescriptions are also combined although not segmented by private or out-of-pocket; those not covered by the province are designated as "non-adjudicated". This represents a major advantage over most of the other provinces where there is no universal drug program. However, this advantage is offset by restricted access policies, which are described below.

In summary, the majority of the Canadian population resides in provinces where the public drug program is restricted to specific segments of the population (e.g., elderly, welfare). Coverage features for the remainder may affect the generalizability of findings obtained in pharmacoepidemiologic studies, and this issue should be addressed in studies on a case by case basis. The population covered by a provincial prescription drug plan is therefore a major criterion for the selection of a Canadian database to conduct pharmacoepidemiologic research.

Prescription drug databases record all prescription drug dispensings received in an outpatient setting. Drugs obtained over-the-counter, in hospital, or in long-term care units are not usually included in the database. Drugs dispensed to nursing home residents are also included if the pharmacy where they acquire their prescriptions is considered to be community-based as opposed to institution-based. Claims databases require that the drug be approved in the formulary before they can be included in the database. Reimbursement is under the jurisdiction of each province. Consequently, the date of inclusion of the drugs in the formulary, and the type of listing (general or restricted), differs from the date of drug approval by Health Canada, and may vary across provinces. This is another important criterion for the selection of a Canadian database.

The main data elements found in the prescription drug databases are listed in Table 17.1. Apart from a few exceptions, data and coding systems are very similar across provinces. With the exception of the Saskatchewan database,[1] prescribed daily dose may be derived directly from the quantity dispensed, prescribed duration, and dose per unit (strength). The prescribed duration may, however, be inaccurate for drugs taken as needed (PRN).

Table 17.1 Information recorded in the prescription drug databases

Common to all provincial databases	Specific to individual databases
Patient information	
Encrypted patient identifier	Category of membership (e.g., welfare recipient, elderly, level of deductible as a proxy for income)
Gender	
	Age: date of birth, birth year, age, or age group depending on database and confidentiality procedures
Drug information	
Date of dispensing	
Drug class (AHFS* classification)	
Drug Information Number (DIN)†	
Generic name	
Brand name	
Strength	
Form of administration	
Quantity dispensed	
Prescribed duration‡	
Prescriber information	
Encrypted prescriber information	
Cost information	
Unit cost	
Patient contribution	
Drug plan contribution	
Total cost, including dispensing fee	

* American Hospital Formulary Service.
† Assigned by Health Canada.
‡ Not available in Saskatchewan.

Indication for a drug prescription is not recorded in any of the databases. For each patient, the year of entry and exit from the drug program are available in the patient information 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.

For each province, the nature and size of the population covered by the public drug program, the database custodians, as well as their year of availability are summarized in Table 17.2. Year of availability refers to the earliest date when data became available through the database custodians, which may not correspond to the year of implementation of the public drug program. Furthermore, depending on the structure of the repository and archive processes, data availability for earlier years may be restricted. Overall, databases may be available for pharmacoepidemiologic research in seven provinces. The three remaining provinces, all located in the East, account for the smallest segments of the Canadian population and have not been used for pharmacoepidemiologic research.

It can be observed from Table 17.2 that approximately half of the databases are accessible through custodians located in a university setting, while the

Table 17.2 Characteristics of the population covered in the public drug database, by province

Province	Total population	Custodian	Population covered	Segments covered			Year of availability
				Welfare	Elderly	Other	
Prince Edward Island	135 000						N/A
New Brunswick	750 000						N/A
Newfoundland/ Labrador	100 000	DHCS*	N/A	✓	✓	Partial	2007
Nova Scotia	1 million	PHRU†	150 000	✓	✓	Partial	
Quebec	7.5 million	RAMQ‡	3.3 million	✓	✓	Partial	1997
Ontario	13 million	ICES§	1.5 million	✓	✓	None	1990
Manitoba	800 000	MCHP**	800 000	✓	✓	✓	2005
Saskatchewan	1 million	SK Health††	910 000	✓	✓	✓	1976
Alberta	3.5 million	AHW‡‡	374 000	✓	✓	✓	Restricted
British Columbia	4 million	PopDataBC§§		✓	✓	✓	1985
		PharmaNet***				✓	(expanded 2003)

* DHCS: Department of Health and Community Services (government).
† PHRU: Population Health Research Unit (university).
‡ RAMQ: Régie de l'assurance-maladie du Québec (government).
§ ICES: Institute for Clinical Evaluative Sciences (non-profit).
** MCHP: Manitoba Centre for Health Policy (university).
†† SK Health: Saskatchewan Health (government).
‡‡ AHW: Alberta Health and Wellness (government).
§§ Population Data BC (university) (formerly available through the British Columbia Linked Health Database (BCLHD).
*** BC Ministry of Health (government).

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.

Medical services databases

The health-care system of Canada is considered to be universal since it covers medical care (although not necessarily prescription drug coverage) for all residents regardless of age and income. For the administration of the program, each province has created a database which includes all claims submitted by the physicians who are paid on a fee-for-service. All patient encounters are individually recorded in the database whether they are provided in an inpatient, outpatient, or emergency department setting. Data elements that are included in each claim and which are relevant for pharmacoepidemiologic research are summarized in Table 17.3.

The nature of the information in the various medical services databases is similar. However, differences exist either in the coding systems or in the categories provided for confidentiality reasons (e.g., physician year of graduation). Practice region is either recorded in the database as a data element, such as in Saskatchewan, or must be derived indirectly from the main practice setting and region identified in a random sample of medical visits for a given physician.[14]

Diagnostic coding depends on the province, following either the International Classification of Diseases, tenth edition (ICD-10) since 2006, or the ICD-9-CM for those provinces that have not yet implemented ICD-10 (e.g., Quebec). Diagnosis is the only field that is not mandatory for payment, which, as shown in the methodologic considerations section below, may pose a threat in the validity of study findings. Procedures are coded according to the Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures.[15] 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 also recorded in a distinct database.[16]

Table 17.3 Information available in the medical services databases

Common to all provincial databases	Specific to individual databases
Patient information	
Encrypted patient identifier	
Gender	
	Age: Date of birth, birth year, age, or age group depending on database
Service information	
Date of encounter	
Service rendered	Coding systems differ across provinces
Location of service (hospital, community clinic, emergency department, long-term care unit, etc.)	Coding systems differ across provinces
Diagnosis	ICD-9 or ICD-10 depending on provinces
Physician information	
Encrypted physician information	
Practice region (urban, rural)	Category and source of information differs according to province
Year of graduation	Categories differ according to province

Hospitalization databases

Unlike the medical services databases, hospitalization databases have been created for the generation 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), and day surgeries. With the exception of Quebec, all provinces contribute to the Discharge Abstract Database (DAD) maintained by the Canadian Institute for Health Information (CIHI).[17] The information is therefore homogeneous across provinces. In Quebec, the hospital discharge database is called Med-Echo.[18]

Table 17.4 Information available in the hospital discharge databases

Patient information
Encrypted identifier
Gender
Age
Hospital admission information
Main diagnosis
Secondary diagnoses
Accident code
Admission date
Discharge date
Length of stay
Hospital identification
Patient destination (community, other hospital, long-term care unit, death)

The information included in the hospitalization databases is summarized in Table 17.4.

In the hospitalization databases, diagnosis was coded with ICD-9-CM until 31 March, 2006, and with ICD-10 ever since. In the DAD database, information on mental-health resources, cancer staging, and reproductive history have been added in 2009–2010. Unlike claims databases, the hospitalization databases are populated once a year and the fiscal year runs from April 1st to March 31st. Databases are typically available 6 months after the end of the fiscal year, at the earliest. In studies where such delay is unacceptable, it is possible to identify hospitalizations through the location of service in the medical services database (physician billings). However, since diagnoses in the hospital discharge database (one principal diagnosis and up to ten secondary diagnoses) are abstracted from hospital charts by medical archivists, they are believed to be more reliable than the single diagnosis recorded by physicians on their billings.

Linkage capacities

The medical services, hospitalization, and prescription drug databases may be linked through a unique identifier, which remains unchanged over time. Linkage provides a longitudinal accumulation of population-based information on all health-care services received by residents of each province. Availability of prescription data may, however, be interrupted by out-of-province residency, admission to acute or long-term care units (no drug data available), or periods of non-eligibility in the public prescription drug programs due to access to private insurance programs, for example. Periods of non-availability of prescription data should be taken into account in the conduct of studies, for example through use of eligibility criteria.

These databases may be linked to other data sources, such as registries (see Chapter 21) or surveys. Technically, linkage with multiple data sources is feasible provided that the health insurance number has been collected. Cancer or infectious disease registries are available in each province, but many database custodians have not yet developed the linkage capacities with these other sources. An exception is Saskatchewan, where many studies involving linkage between the cancer registry and claims databases have been conducted.[1] Some provincial database custodians, such as those of Ontario, Nova Scotia, and Manitoba, also act as a warehouse for other databases. Consequently, the linkage processes are already in place in those provinces.

Linkage capacities include linkage with other statistics or health databases, registries (e.g., cancer registries, vital statistics), national health surveys, or with database/ registries created for research or clinical purposes (e.g., Dalhousie Multiple Sclerosis Research Unit database,[19] Canadian Cardiac Network in Ontario[20]). Although many databases have been identified across Canada, many have not yet been linked to health-care databases for the purpose of pharmacoepidemiologic research.

Database access and confidentiality

Access policies are a very important consideration in the selection of a database. As described below, some are available to academic researchers only, while others may be accessible by employees of the pharmaceutical industry also. Access may be sought directly from government agencies in Saskatchewan, Quebec, and British Columbia. For the other provinces, access is sought through a repository, which may be an academic center (e.g., Nova Scotia,

Ontario) or through a regional health board (e.g., Manitoba). For seven provinces (SK, QC, ON, BC, MB, NS, AB), health databases have been widely used for pharmacoepidemiologic research. For the remaining provinces, access is restricted because of absence of a common database custodian and/or absence of access procedures. Furthermore, databases from British Columbia, Manitoba, Newfoundland/ Labrador, and Ontario are not accessible by the pharmaceutical industry. In Ontario and Manitoba, access is restricted to designated researchers only. The databases from Saskatchewan and Quebec are accessible to all researchers, regardless of sector.

Regardless of the province, a database request must be submitted to the custodians for review. Review consists of ethics approval and, for university-based custodians, a scientific review as well.

The data sets available to researchers vary greatly across provinces. In Saskatchewan, Quebec, and Nova Scotia, raw anonymized datasets are sent directly to the researchers. In Ontario, data must be analyzed in-house by a member or affiliate of the Institute for Clinical Evaluative Sciences (ICES). To maintain confidentiality of the data, no patient, health-care professional (including pharmacist), or institution identifiers are transmitted to researchers. All identifiers are encrypted. Furthermore, in Quebec, to reduce the possibility of identifying a particular patient, only a random sample of up to 125 000 of the population eligible for a given study may be obtained, and no birthdates are transmitted. Patient age is categorized in 5-year intervals. Year of death may be obtained but exact date of death is not provided. The latter would be available through the vital statistics database. More detailed information may be sought, but a special request must be submitted to the provincial Information Privacy Commissioner, which adds considerable delay to the process. In Saskatchewan, all drugs that are part of the data elements required for a given study must be identified *a priori.* Data extraction does not offer the possibility to obtain all drugs acquired by a given patient.

The time required for database extraction also varies across provinces, ranging from 10–20 weeks to 1 year. Timelines are usually shorter for studies conducted directly by database custodians than by external researchers. Ethics approval may also add a delay when the ethics committee is not located within the database custodian organization. For example, in Quebec, approval from the Information Privacy Commissioner must be sought prior to data extraction, which adds considerable delay if sensitive data are requested (e.g., birth date versus age group). There are also great differences in costs of extraction across provinces. A recent survey, which involved contacting database custodians with a case study, confirmed that predicted delay of access includes a backlog of data requests and access approval, and ranges from 3–4 months (Alberta, Quebec) to approximately 12 months (Saskatchewan).[21] In practice however, delays may be greater and are highly variable from one request to another since they depend on the workload. For example, requests made directly to custodians that are government agencies typically take longer, since the task is in addition to other administrative requests, and the agencies are not dedicated exclusively to research. For example, over a 1-year period, the Quebec database custodian (Régie de l'assurance-maladie du Québec, RAMQ) received 1400 data requests by researchers and other stakeholders in the health-care system.[22]

Linkage with medical charts or complementary sources of information

The variable validity of diagnostic codes recorded in health-care databases is a well-known threat to pharmacoepidemiologic studies. In many instances, it may be desirable to access source information, such as hospital or outpatient charts, in order to obtain clinical characteristics and validate the diagnosis of cases ascertained in the databases. To obtain claims data on patients identified in a clinical or hospital setting is feasible, mainly through informed consent. However, one major barrier for the conduct of validation studies is the feasibility to link the information that appears in the claims databases back to the individual patient charts for diagnosis confirmation. Because of data protection rules and regulations, this process requires approval from the provincial information access

commissioner and, in some provinces, it may not be acceptable.

Yet, some validation studies have been conducted and may be found in the literature, but validation data are far from comprehensive. According to a recent review,[23] at least 18 validation studies of Canadian databases have been published in the literature, the majority of which having been conducted to validate the diagnostic codes present in the medical services databases. Because the method of ascertainment of cases of adverse events in the database, as well as the condition under study, are very heterogeneous across validation studies, a quantitative synthesis of findings is not warranted. Furthermore, the validity of the diagnostic codes used in medical claims databases may also vary across provinces, as they might depend on the pattern of health-care delivery, and hence billing practices. This hypothesis has not yet been substantiated in the literature.

Another methodologic issue that would warrant access to patients or their health records is the collection of data that are not present in the databases and that may be confounders. Indication, smoking, alcohol use, body mass index, and over-the-counter drug use are examples of such potential unmeasured confounders. Two-stage sampling is a method that may be used to address unmeasured confounders through the collection of supplementary data in a subset of individuals. This requires identifying patients in the claims database and using nominal information to access medical charts.[24] To our knowledge, this has been done only through the Saskatchewan[24] and Quebec databases.[25]

Strengths

Canadian databases have been widely used for pharmacoepidemiologic research. Although most studies found in the literature involve a limited number of databases, such as those from Saskatchewan, Ontario, Quebec, British Columbia, and Manitoba, as research capacity within Canada is expanding, one can foresee greater use of the other databases as well. Among the unique features of the Canadian databases are the availability of longitudinal population-based data on prescription drugs and health-care use, and most importantly linkage capacities with other sources of data, such as registries or surveys.

Weaknesses

The technical advantages of the Canadian databases are somewhat offset by decentralization and the access restrictions described above. Although efforts to pool databases are currently ongoing in a number of countries, few attempts have been made in Canada to date. The difficulty in accessing patient health records through nominal information, for diagnostic validation or to control for unmeasured confounders, is a threat to the validity of studies that aim at assessing the safety or effectiveness of a drug.

Particular applications

Below are a few examples of studies conducted through linkage with various databases. Examples illustrate drug utilization studies, risk evaluation studies, and comparative effectiveness research.

Drug utilization

Dalhousie Multiple Sclerosis Research Unit (DMSRU) database: the DMSRU database includes 25 years of clinical data and may be linked to Nova Scotia's provincial data. In a study conducted by Sketris *et al.*,[19] drug use and costs by seniors with multiple sclerosis was compared to that of all senior residents of the province.

Through the *National Rehabilitation Reporting System* of Ontario, a study was conducted on health-care utilization in non-traumatic and traumatic spinal cord injury patients using linkage with the *Ontario Health Insurance Plan* for physicians' fee-for-service claims and the *National Ambulatory Care Resource System* for all visits to emergency departments.[26]

Another study compared health-care use before and after a workplace injury using linkage

between the *Workers' Compensation Board (WCB)*[27] of British Columbia and the *British Columbia Linked Health Database*.[28] All these databases may be accessible through PopulationData BC. Medical services and hospitalizations were obtained for 5 years before and 5 years after the injury. Use was also compared to injured workers who did not file a claim.

Comparative effectiveness research

A non-randomized comparative effectiveness study on drug-eluting stents impregnated with paclitaxel versus sirolimus in diabetic and non-diabetic patients has also been conducted through linkage between patients identified in the *Cardiac Care Network of Ontario Percutaneous Coronary Intervention Registry* and linked to the administrative health databases of Ontario.[29] Study outcomes consisted of target-vessel revascularization, myocardial infarction, and death.

Risk evaluation studies

In Manitoba, there is a universal *Manitoba Bone Density Program*, which is accessible to all residents of the province provided that they meet the criteria for testing consistent with published guidelines.[30,31] The program has managed all clinical bone density scanning (dual-energy X-ray absorptiometry testing) of the province since 1997, and the database has been shown to be over 99% complete and accurate.[32] A study conducted by Morin *et al.*[33] assessed the association between weight and body mass index, and low bone mineral density and fractures in women age 40 and 59 years.

A study on the association between benzodiazepines and motor vehicle accidents also involved a specific linkage between the *Quebec driver's license files*, police reports of injurious crashes, and healthcare databases.[34]

Canada-wide linkage has rarely been done, and never with prescription drug databases since they are province-specific. However, the *Canadian Organ Replacement Register (CORR)* is a national organ failure registry in Canada and is maintained by the Canadian Institute for Health Information (CIHI).[35] This database was linked with the CIHI hospital discharge database which is also centralized. Such a linkage allowed researchers to create the Canadian Pediatric End-Stage Renal Disease database, a tool to perform longitudinal studies in this patient population.[36] This database includes all provinces except Quebec, which has separate privacy legislation.

Other linkage capacities include health-care databases with population health surveys,[37] Provincial Vital Statistics databases (birth, death, and cause of death), Canadian Reduction of Atherothrombosis for Continued Health (REACH) Registry,[38] Nova Scotia, Saskatchewan, British Columbia Cancer Registry (available through database custodians), and Ontario diabetes database. According to Jacobs and Yim,[39] there are many such clinical databases in Canada and they will be increasingly linked, but so far it has been uncommon.

The future

Medical services, hospitalizations, and prescription drug databases are widely available in Canada. At present, their potential for use in pharmacoepidemiologic research is somewhat hampered by access restriction and/or delays. Great potential exist for linkage with complementary sources of information such as registries. However, such registries are often local initiatives, with no information centralized in a common repository. Greater linkage capacities between databases and other complementary data sources would be valuable to augment the data sources in Canada, such as laboratory test results. At the present time, the Manitoba bone mass registry is the only provincial source of data on diagnostic test results.

The fragmentation of databases across provinces and heterogeneity in custodians leads to a taxonomy of databases that is very complex. In the future, database access and linkage capacities will need to be better communicated in order to implement collaborative projects across provinces.

Acknowledgements

The authors would like to thank the Office of Pharmaceuticals Management Strategies of Health

Canada for providing us with the opportunity to conduct the review of "Data sources to support research on real world drug safety and effectiveness in Canada". *The distribution of the Report by Health Canada shall not be construed as constituting an endorsement by Health Canada, or any other notice approved in writing by Health Canada.* We are extremely grateful to all the researchers and key informants from each province who have provided us with valuable and practical information on the availability of the databases. Such support is certainly a stepping stone for the success of collaborative research involving databases that will be implemented in Canada in the context of the drug regulation reform.

References

1. Downey W, Beck P, McNutt M, Stang M, Osei W, Nichol JL. Health databases in Saskatchewan. In: Strom BL, ed. *Pharmacoepidemiology*, 3rd edn. John Wiley & Sons Ltd, 2000.
2. Moride Y, Metge C. *Data Sources to Support Research on Real World Drug Safety and Effectiveness in Canada: an Environmental Scan and Evaluation of Existing Data Elements*, 2010. http://www.pharmacoepi.ca (accessed Feb. 2011).
3. Yeates N, Lee DK, Maher M. Health Canada's progressive licensing framework. *Can Med Assoc J* 2007; **176**: 1845–7.
4. Canadian Institutes for Health Research. *The Drug Safety and Effectiveness Network (DSEN)*. http://www.cihr-irsc.gc.ca/e/39389.html (accessed Oct. 2010).
5. Ontario Ministry of Health and Long-term Care. *Ontario Drug Benefit Program.* http://www.health.gov.on.ca (accessed Oct. 2010).
6. Régie de l'assurance-maladie du Québec. http://www.ramq.gouv.qc.ca/fr/statistiques/banques/vuedensemble.shtml (accessed October 2010).
7. Government of Alberta. Health and Wellness. *Prescription Drug Program*. http://www.health.alberta.ca/AHCIP/prescription-program.html (accessed Oct. 2010).
8. BC Ministry of Health. http://www.health.gov.bc.ca/pharmacare/ (accessed Oct. 2010).
9. Newfoundland and Labrador. Department of Health and Community Services. *Prescription Drug Program*. http://www.health.gov.nl.ca/health/prescription/index.html (accessed Oct. 2010).
10. Nova Scotia Ministry of Health. *Nova Scotia Pharmacare*. http://www.gov.ns.ca/health/pharmacare/ (accessed Oct. 2010).
11. Manitoba Health. *Manitoba Pharmacare Program*. http://www.gov.mb.ca/health/pharmacare/ (accessed Oct. 2010).
12. Saskatchewan Health. *Drug Plan.* http://www.health.gov.sk.ca/drug-plan-benefits (accessed Oct. 2010).
13. BC Ministry of Health. *PharmaNet*. http://www.health.gov.bc.ca/pharmacare/pharmanet/netindex.html (accessed Oct. 2010).
14. Moride Y, Du Fort GG, Monette J, Ducruet T, Boivin JF, Champoux N, *et al*. Suboptimal duration of antidepressant treatments in the older ambulatory population of Quebec: association with selected physician characteristics. *J Am Geriatr Soc* 2002; **50**: 1365–71.
15. Statistics Canada. *Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures*. http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=82-562-X&lang=eng (accessed Oct. 2010).
16. Population Data BC. *Data Set: Mental Health*. http://www.popdata.bc.ca/data/internal/health/mentalhealth (accessed Oct. 2010).
17. Canadian Institute for Health Information. http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=services_dad_e (accessed Oct. 2010).
18. Quebec Ministry of Health. *Portal of Occupational Health and Safety (OHS) Data Bases and Measurement Instruments*. http://www.portail-rrsstq.com/fiches/Med_echo.html (accessed Oct. 2010).
19. Sketris IS, Brown M, Murray TJ, Fisk JD, McLean K. Drug therapy in multiple sclerosis: a study of Nova Scotia senior citizens. *Clin Ther* 1996; **18**: 303–18.
20. Canadian Cardiac Network of Ontario. http://www.ccn.on.ca/ (accessed Oct. 2010).
21. Rawson NS. Access to linked administrative healthcare utilization data for pharmacoepidemiology and pharmacoeconomics research in Canada: anti-viral drugs as an example. *Pharmacoepidemiol Drug Saf* 2009; **18**: 1072–80.
22. Régie de L'assurance-Maladie du Québec (RAMQ). https://www.prod.ramq.gouv.qc.ca/IST/CD/CDF_DifsnInfoStats/CDF1_CnsulInfoStatsCNC_iut/DifsnInfoStats.aspx?ETAPE_COUR=2&LANGUE=en-CA (accessed June 2011).
23. 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 Safety* 2009; **18**: S1–S273.

24. 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**: 309–15.
25. Breton MC, Beauchesne MF, Lemière C, Rey E, Forget A, Blais L. Risk of perinatal mortality associated with asthma during pregnancy: a 2-stage sampling cohort study. *Ann Allergy Immunol* 2010; **105**: 211–7.
26. Guilcher SJT, Munce SEP, Couris CM, Fung K, Craven BC, Verrier M, *et al*. Health care utilization in non-traumatic and traumatic spinal cord injury: A population-based study. *Spinal Cord* 2010; **48**: 45–50.
27. BC Workers' Compensation Board. http://www.popdata.bc.ca/data/internal/occupational (accessed Oct. 2010).
28. Brown JA, McGonough P, Mustard CA, Shannon HS. Health care use before and after a workplace injury in British Columbia, Canada. *Occup Environ Med* 2006; **63**: 396–403.
29. Chiu M, Ko DT, Austin PC, Cohen EA, Velianou JL, Goeree R, *et al*. Paclitaxel versus sirolimus stents in diabetic and nondiabetic patients. *Circ Cardiovasc Qual Outcomes* 2009; **2**: 96–107.
30. Manitoba Health. *Manitoba Bone Density Program*. http://www.gov.mb.ca/health/bonedensity/index.html (accessed Oct. 2010).
31. Leslie WD, Metge C. Establishing a regional bone density program: lessons from the Manitoba experience. *J Clin Densitom* 2003; **6**: 275–82.
32. Leslie WE, Caetano PA, Macwilliam LR, Finlayson GS. Construction and validation of a population-based bone densitometry database. *J Clin Densitom* 2005; **8**: 25–30.
33. Morin S, Tsang JF, Leslie WD. Weight and body mass index predict bone mineral density and fractures in women aged 40 to 59 years. *Osteoporos Int* 2009; **20**: 363–70.
34. 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**: 27–31.
35. Canadian Institute de Health Information. *Health Services Databases*. http://www.cihi.ca (accessed Oct. 2010).
36. Samuel SM, Tonelli MA, Foster BJ, Nettel-Aguirre A, Williams R, Soo A, *et al*. Overview of the Canadian pediatric end-stage renal disease database. *BMC Nephrology* 2010; **11**: 21–6.
37. 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.
38. Smolderen KG, Bell A, Lei Y, Cohen EA, Steg PG, Bhatt DL, *et al*: REACH registry investigators. One-year costs associated with cardiovascular disease in Canada: Insights from the REduction of Atherothrombosis for Continued Health (REACH) registry. *Can J Cardiol* 2010; **26**: 297–305.
39. Jacobs P, Yim R. *Using Canadian Administrative Databases to Derive Economic Data for Health Technology Assessments*. Ottawa ON: Canadian Agency for Drugs and Technologies in Health, 2009.

CHAPTER 18
Pharmacy-based Medical Record Linkage Systems

Ron M.C. Herings[1,2] *and Lars Pedersen*[3]
[1] PHARMO Institute, Utrecht, the Netherlands
[2] Department of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, the Netherlands
[3] Department of Clinical Epidemiology, Aarhus University Hospital, Denmark

Introduction

In this chapter we discuss pharmacy-based medical record linkage systems, which each have a large pharmacy-based dispensing database as its primary exposure database. This database is linked to a central patient router file, which is subsequently linked to outcomes databases (e.g., hospitalization, mortality, clinical laboratory), of which some are owned and governed by different organizations. The central patient router file includes unique, anonymized patient identification numbers, which are used to bring together defined exposure and outcome data on patients stored in the different linked databases, into a single new research database. This combined research database is then used to perform pharmacoepidemiologic studies.

Pharmacy-based medical record linkage systems are best described as federated or virtual database networks which transparently integrate multiple autonomous databases into a single system in which data are linked using anonymized patient identification numbers. Since the constituent databases remain autonomous, these systems are an alternative when integrated databases are not available, or where stringent privacy and governance rules do not permit storage of data in a central repository. Access to the individual databases to build a new, linked dataset for scientific research then requires permission of the individual organizations. In such schemes individual organizations, which may vary in size from a single general practitioner (GP) to a large national database, have maximum control to comply with their own governance rules. In Denmark, researchers can link several national registries using unique patient identification numbers provided that the registries and laws permit such linkage. In time, these networks may partially 'de-federate' if organizations involved permit storage of copies or subsets (e.g., regional) of the autonomous databases into a new repository. For instance, two or more autonomous organizations can decide to store overlapping regional subsets of their databases in a new database to perform pharmacoepidemiologic studies. In the Netherlands, hospitals, clinical laboratories, GPs, and pharmacists store subsets of their databases to be linked for pharmacoepidemiologic research in the PHARMO (PHARmacoMOrbidity linkage system) Institute.

Pharmacy-based medical record linkage systems have been established in the Nordic countries (Denmark, Sweden, Norway, Finland, Greenland, and Iceland),[1–5] Scotland,[6] and the Netherlands.[7] In these countries, prescriptions issued by both GPs and medical specialists are filled in community pharmacies. The information in these community pharmacy databases can be combined into a single national

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

database (e.g., the Danish National Prescription register) or by some organizations into regional databases (e.g., the Dutch PHARMO system). These exposure databases are then linked either with national registries or overlapping regional databases. Some of these countries' pharmacy-based medical record linkage systems are among the oldest systems for pharmacoepidemiologic research in Europe, created in the early 1990s, and have been used as resources for hundreds of publications. Comparison among these record linkage databases is highly complex because of country-specific differences in laws and regulations that affect content, completeness, and validity of the linked, autonomous databases. Moreover, in most EU countries, many specific record linkage based systems are available, ranging from several thousand up to millions of patients, but are limited to a particular disease and have limited use for pharmacoepidemiologic research. The pharmacy-based medical record linkage systems in the Nordic countries and the Netherlands are remarkably similar with respect to available exposure data (Table 18.1) and linked clinical outcome data (Table 18.2). Space restrictions preclude a

Table 18.1 Data available in drug exposure files in Denmark, Northern countries, and the Netherlands

Record linkage system	OPED* (1990)	AUPD† (1989)	Other northern countries	PHARMO‡ (1986)
Geographic area	Regional	Regional	National	Regional
Population	1.2 million	1.7 million	17 million	3.2 million
Pharmacy				
Unique identifier	YES	YES	YES	YES
Location	YES	YES	YES	YES
Monthly updated	YES	YES	YES	YES
Dispensed drugs				
Unique identifier	YES	YES	YES	YES
ATC code	YES	YES	YES	YES
DDD number	YES	YES	MOST	YES
Amount dispensed	YES	YES	YES	YES
Prescribed dose	NO	NO	FREETEXT	YES
Reimbursed drugs	YES	YES	SOME	YES
Non-reimbursed drugs	NO	NO	MOST	YES
Duration of use	NO	NO	SOME	YES
Dispensing date	YES	YES	YES	YES
Indication for use	NO	NO	NO	NO
Prescriber				
Unique identifier	YES§	YES	SOME	YES
Profession**	YES	YES	YES	YES
Practice	YES	YES	SOME	YES
Date started practice	YES	YES	SOME	YES
Year of birth	YES	YES	YES	YES
Sex	YES	YES	YES	YES

* Odense University Pharmacoepidemiologic Database.
† Aarhus University Prescription Database, Denmark.
‡ PHARMO Record Linkage Network.
§ PHARMO ID, after de-duplication.
** Physician, nurse, dentist, midwife.

Table 18.2 Linked databases in the Danish network and the PHARMO record linkage system

Database	Characteristics	
	Denmark	PHARMO
Clinical laboratories http://www.pharmo.nl/	Test name, IUPAC test code, local analysis number, result, measurement unit, date or ordering and carrying out the analysis Different laboratories information system Population subset 1999–2009	Test name, WCIA-test code, local analysis number, result, measurement unit, date or ordering and carrying out the analysis PHARMO ClinLab Population subset (1.2 million) 1991–2010 (update 3 month)
Birth registers	Multiplicity (singleton, twin etc.), weight, length, fetal presentation, gestational age, Apgar scores, congenital disease, mode of delivery Maternal status includes previous stillbirths, live birth (parity), age at delivery, smoking, location of birth Mortality Danish Medical Birth Register http://www.sst.dk	Multiplicity (singleton, twin etc.), weight, length, fetal presentation, gestational age, Apgar scores, congenital disease, mode of delivery Maternal status includes previous stillbirths, live birth (parity), age at delivery, smoking, location of birth Mortality Dutch Perinatal Registration http://www.perinatreg.nl
Hospitalizations	Admission/ discharge date, diagnoses (ICD-10), operations, surgeries (ICD-10) and selected in-hospital treatments Danish National Registry of Patients http://www.sst.dk	Admission/ discharge date, diagnoses (ICD-10), operations and surgeries (ICD-10) IUPAC testcode, local analysis number, result, measurement unit, date or ordering and carrying out the analysis, ID hospital or GP ordering the test Dutch Hospital Data http://www.dutchhospitaldata.nl/
Death Registry	Date of death, cause of death Danish Registry of Causes of Death http://www.sst.dk	Date of death, demographic history National Centre Family history http://www.cbg.nl/
Cancer	Date of cancer diagnosis, method of verification, morphology, topography, initial treatment, surgery, radiotherapy, chemotherapy, hormonal and immunotherapy, co-morbidity Additional information for specific cancers http://www.sst.dk	Date of cancer diagnosis, method of verification, morphology, topography, initial treatment, surgery, radiotherapy, chemotherapy, hormonal and immunotherapy, co-morbidity at diagnosis Additional information for specific cancers http://www.ikcnet.nl/
Pathology	Test date, pathological specimens, morphology, topography procedures, diagnoses National Pathology Registration http://www.sst.dk	Test date, pathological specimens morphology, topography, procedures, diagnoses PALGA http://www.palga.nl

detailed description of all databases that are part of these schemes. More detail can be found in the websites of the individual record linkage systems.[8,9]

Therefore, as examples of pharmacy-based medical record linkage systems, we will discuss the Danish OPED (Odense University Pharmacoepidemiologic Database), the Danish AUHD (Aarhus University Prescription Database;[10,11] AUHD is also referred to in published literature by the names Pharmacoepidemiologic Prescription Database of North Jutland and Prescription Databases of North Denmark and Central Denmark Regions), and the Dutch PHARMO record linkage system.[9]

Description

Pharmacy-based medical record linkage systems have at least three types of files in common: a patient router file containing the characteristics of patients in the catchment population; a pharmacy-based drug exposure registry; and one or more linked clinical registries obtained from other organizations or health professionals. The exposure and clinical registries are linked to the patient router file using a variety of record linkage methods. Below, we first discuss the methods used to link the different databases. Then we discuss the linked outcomes databases.

Record linkage

At least three methods are used to link different patient-based databases.[12–14] The most straightforward is a *deterministic* linkage, based on unique personal identification numbers used across multiple systems. This is the main linkage method for the OPED and AUPD databases in Denmark and can also be used for the entire population of Denmark.[5] Alternatively, in the absence of unique personal identification numbers, a sequence of patient characteristics can be used to construct a unique, semideterministic linkage key. The Medicine Monitoring Unit (MEMO) database[15] in the Tayside region of Scotland uses such a key—the Community Health Index (CHI) number. This is a ten-digit integer, in which the first six digits indicate the date of birth, digits seven and eight provide information on region of residence, the ninth digit indicates sex, and the tenth digit incorporates a checksum to ensure the number's validity. A third method, probabilistic record linkage, is used to link databases and registries in the PHARMO system.[7,16] It relies on several patient characteristics to create identifiers. Records from two files to be linked are grouped into record pairs, based on initial agreement of date of birth and gender. Additional characteristics then are added to each side of the pair, such as first initial of surname, surname, postal code, name of general practitioner, name compression algorithms (e.g., Soundex code),[17] and date of death. Bayesian likelihood estimations and learning-based rules are applied to estimate the likelihood that two records from two distinct files belong or do not belong to the same individual.[12,13,18,19]

All linkage approaches are susceptible to error. Although the use of personal identification numbers (deterministic or semideterministic) is often regarded the most accurate linkage strategy, such numbers may have been used previously, may change over time, may be recorded incorrectly, or may be missing.[20] Validation of these linkages can be difficult because comparison with name and address in a sample of patients requires stringent protocols and permission. The semideterministic linkage used in the MEMO system yielded sensitivity and positive predictive values of 0.96 and 0.95, respectively,[6] when compared to name and address information in a sample of patients. A recent validation study of the linkage of cancer registrations to the PHARMO roster files confirmed both a sensitivity and specificity of 0.98[7,21] compared to the name and address information of a sample of patients. The PHARMO record linkage system uses a combination of deterministic (e.g., GP data), semideterministic (e.g., clinical laboratory data), and probabilistic (e.g., hospital data) techniques to link databases to a central patient roster file. Probabilistic techniques are also used to identify duplicates in patient files from community pharmacies. Whatever linkage method is applied, a unique patient identification number is available in patient router and event files.

Patient router files

A patient router file, in which all patients have a unique identification number, represents essentially all inhabitants of a population in a defined geographic area, such as a region or country. Patient router files typically include a unique personal identifier, year of birth, gender, and place of residence, and serve as a router pointing to the files where exposure and clinical histories are stored. Patients may enter the population by birth or immigration, and may depart from the population through death or emigration. In Denmark, information regarding date of birth and date of death is available from the Civil Registration System (CRS). In the Netherlands this information is obtained through linkage to birth or death registries. The time periods between date of entry into and date of exit from the catchment area are defined as the "event eligible" periods for individual patients. Within these eligible periods, cohorts can be extracted based on defined events (e.g., exposure, outcomes) or particular patient characteristics (e.g., age). Follow-up almost always ends with death or end of the registration period, if not otherwise specified. Changes in health care in recent years have led patients to shop among community pharmacies, creating the need to identify duplicate patients and to define new unique patient numbers.[16]

Exposure databases

Community pharmacies are the source of exposure information in pharmacy-based medical record linkage systems.[1,10] In both Denmark and the Netherlands, for example, drugs prescribed by GPs, medical specialists, or others with a prescribing license have to be filled in community pharmacies. In response to financial and other incentives, most patients have designated a single pharmacy to fill all their prescriptions. The PHARMO pharmacy database holds dispensing information going back to 1986; OPED began collecting data in 1990 and AUPD in 1989. There are about 1600 community pharmacies in the Netherlands and 300 in Denmark. Prescriptions originating from the primary healthcare sector account for approximately 96% of the total volume of medicinal product sales in Denmark. These databases typically use the Anatomical Therapeutic Chemical (ATC) coding system (see Chapter 24) to code drugs.

However, some important national and regional differences should be noted. In the Danish Government Statistical Office, a nationwide prescription database has been available since the 2004, regardless of reimbursement status, but all analyses must be made on their server, with no access to the civil registration number. Registration in the regional Danish prescription databases depends on reimbursement status, and hence data on sedatives, hypnotics, oral contraceptives, and laxatives are incomplete. The Dutch PHARMO database is not restricted to reimbursed drugs, but also contains information on non-reimbursed drugs, homeopathic drugs, and herbal remedies, as well as some medical devices (e.g., blood glucose monitors), urinary incontinence pads, and other non-pharmaceutical products. The PHARMO database also includes dosage instructions and well as the legend defining duration of use. Recording of the exact legend for duration of use is important for pharmacists' medication surveillance routines. No database includes the reason for prescribing (i.e., therapeutic indication for the drug) in the dispensing records. However, in the PHARMO record linkage network, the indication for a sample of dispensed drugs can be ascertained or inferred through linkage to medical records.

The PHARMO Institute also holds a database of in-hospital drug exposures since 2000 from 12 out of 80 hospital pharmacies in the Netherlands, within a total catchment population of 2.5 million. The 12 hospitals record all medication orders for patients during their hospital stays. Data for approximately 1 million of these 2.5 million hospitalized patients are linked to the PHARMO patient router files. Coding problems may occur with in-hospital drugs that are compounded specifically for individual patients (mainly cancer treatments), requiring extra coding via chart abstraction in hospital pharmacies. Inpatient medication data are used primarily to identify drugs dispensed to treat cancer and infectious diseases, and to study differences between hospital-based and community drug use.[21] A similar population database is being

established in Denmark using electronic patient files. Currently, it is limited to a few drug products. However, since 1999, the most expensive inpatient treatments (for instance, biological treatments) and some intravenous treatments have been recorded in the Danish National Registry of Patients. Hospital databases present many challenges, especially with respect to identification and coding of dispensed products.

Hospitalizations

Denmark has a registry containing information on hospitalized patients since 1977 and the Netherlands has had a similar registry since 1963. They completely cover the period for which drug exposure information is available, and include all clinical discharge diagnoses. In Denmark, diagnoses are coded using either the *International Classification of Diseases, 8th edition* (ICD-8) or 10th edition (ICD-10) systems. In the PHARMO region, discharge diagnoses are coded using ICD-9-CM (clinical modification). Denmark's hospital registry also includes information on visits to emergency departments and outpatient hospital clinics since 1995. In the Netherlands, the hospital registry will be converted to a new database over the next few years, coded according to the ICD-10 system, and also will include information on visits to outpatient clinics and emergency units. Both the Denmark and Dutch hospital databases include information on more than ten discharge diagnoses and procedures per hospitalization, but are incomplete regarding non-invasive diagnostic procedures such as magnetic resonance imaging and computed tomography. In Denmark and the Netherlands, physicians code medical treatments and surgeons code surgical procedures. Trained coding specialists in the hospitals check the data and enter them into the system. Data are updated and uploaded from the hospitals to the national registries monthly in Denmark and annually in the Netherlands and are available after control procedures are completed.

Clinical laboratory data

Both the PHARMO record linkage system and the Danish databases have access to laboratory results. Data are collected from individual laboratories, including all clinical tests ordered by GPs and hospitals. Orders for tests and test results are communicated electronically between GPs and hospitals. Individual laboratories in Northern Denmark and in the PHARMO region use different software systems. In the Netherlands data are standardized, coded, and cleaned in the PHARMO Institute, and in Denmark this process takes place in the Department of Clinical Epidemiology, Aarhus University Hospital. For each study, permission must be obtained from the PHARMO network or Aarhus University Hospital to access the databases.

The PHARMO database includes data on more than 500 different hematological and serological tests for more than 2.3 million patients, of whom 1.2 million overlap and are linked to the PHARMO roster files. Data are updated every 3 months. In Denmark, the population covered by the laboratory database is located in a well-defined geographical area representing approximately 30% of the population. Data completeness has not been fully assessed for the early years (1990s) of the Danish laboratory database, but it is considered complete from January 1, 1999 through December 31, 2009. A limitation of most laboratory tests is that they are performed because of clinical suspicion of a particular disease. On the other hand, some tests related to diabetes, asthma, chronic obstructive pulmonary disease, and cardiovascular treatment are part of standardized diagnostic treatment regimens introduced in 2006. For these diseases, GPs are reimbursed according to a fixed payment for the condition, and these tests are performed at least once each year for the majority of patients. These tests include lipid and glycosylated hemoglobin (HbA1c) levels in patients with diabetes. Non-laboratory measurements include blood pressure, smoking status, body mass index, and results of fundoscopic examinations. At the PHARMO Institute, some test results are recorded in a function test database (e.g., blood pressure, electrocardiograms [ECGs], forced expiratory volume in 1 second, fundoscopic examinations, and microfilament assessments), and are available for research.

Pathological findings

In both countries, clinical pathology data are transferred by national organizations into a national pathology registry. In Denmark, pathology data have been used for research and measurement of quality in diagnostics and treatment since 1997. In the Netherlands, the Pathological Anatomy National Automated Archive (PALGA) register is a central depository of pathological findings used in routine daily practice. Excerpts of pathological findings are anonymized, and an excerpt of the complete dataset is stored in a National research database, accessible to researchers upon request. The PALGA network includes data from all 64 histopathology and cytopathology laboratories in the Netherlands, with a continuously expanding automated archive of excerpts of pathology reports (currently about 42 million excerpts on nearly 10 million patients since 1991.[22] In both Denmark and the Netherlands, the pathology databases are used to assess the quality of cancer registrations, and to identify and alert cancer registries about new cases. Pathology in Denmark is coded according to the Systemized Nomenclature of Medicine (SNOMED); a classification system comparable in scope to SNOMED is used in the Netherlands.

Cancer registries

In Denmark and the Netherlands, national organizations and governments have established cancer registries to collect data on incident cancers. In both countries, registries are based on notification forms completed at the hospital level. In the Netherlands, trained staff members from one of the regional cancer centers visit hospitals to abstract information from medical records onto specific data forms. The PHARMO record linkage system is restricted to the Eindhoven Cancer Registry (IKZ), which covers a catchment area of approximately 2 million patients; to date 1 million patients are linked to the PHARMO roster files.[21,23] By this linkage, incidence cases, staging, and other therapies are added to the PHARMO record linkage scheme. The data recorded in the Danish and Dutch national cancer registries are comparable, with diagnoses coded using ICD-10. Unlike other Dutch regional cancer centers, the Eindhoven Cancer Registry records include basal cell carcinomas and co-morbidity data at the time of first occurrence of a cancer.

Birth registry, maternal linkage, and death

The National Perinatal Registration (PRN) program in the Netherlands (PHARMO) and the Medical Birth Registry in Denmark were established for surveillance of birth rates and other factors related to birth. Since 1973, the Danish Medical Birth Registry has stored prospectively collected information on all live births and stillbirths, including the personal identification numbers of both infant and mother. Other variables are maternal age, body mass index, maternal smoking status during pregnancy, pregnancy complications, and infant characteristics. Both the PRN database and the Danish Medical Birth Registry are used for research purposes. Linkage of these registries to the core pharmacy databases permits study of drug teratogenicity, delivery complications, and maternal diseases, as well as short-term and long-term consequences to children of *in utero* drug exposure. Another benefit is left-censoring of patient router files. This is particularly important for the PHARMO database because the linkage process indicates whether children are born in the PHARMO catchment area or elsewhere. In the PHARMO system, linkage of maternal drug use and children's health is limited to the time when children live with their parents, but in Denmark all family relations can be identified since 1968. In contrast, use of the civil personal registration number in Danish registries allows very long-term follow-up of infants and mothers. Both the Dutch and Danish databases cover all births delivered at home or in the hospital. Mortality statistics are available and coded in the patient router files through linkage to national vital statistics databases in Denmark, and through linkage to the genealogy database in the Netherlands. The PHARMO system is by law not allowed to link to the registry of cause of death. Both registries allow researchers to study all-cause death, and to censor subjects at the time of death.

General practitioner medical records

The GP database is currently the fastest growing database in the PHARMO Institute. It now includes the medical records of more than 1 million patients since 2004. In structure and content it is similar to the Integrated Primary Care Information (IPCI) database (http://www.ipci.nl). The use of data from the GP database requires permission from participating GPs, specific to each research proposal. The IPCI database and the PHARMO GP database are currently being merged into a single database covering more than 2.5 million patients in general practice, of whom 500 000 are linked to the PHARMO drug database. Data include an International Classification of Primary Care (ICPC)-coded problem list, treatments, drug prescriptions, laboratory tests and other measurements, and referral information to medical specialists. The linkage of the PHARMA GP and IPCI databases permits study of relationships between prescribing and dispensing information, validation of hospitalizations on the basis of anonymized discharge letters, and crosschecking among medical events and the clinical laboratory and function test databases (for example, smoking status, activities of daily living, function tests, indication for prescribing, etc.) to ascertain completeness of data. Diagnoses are coded according to the ICPC system, and laboratory tests and function tests are coded according to the Dutch national standard coding systems (Werkgroep Coordinatie Informatisiering en Automatisiering; WCIA); drugs are coded with national product identifiers and ATC code.

A system of linkable registries comprises the Primary Health Care Database in Denmark, maintained by the National Health Service. It was established to reimburse all health-care services provided by general practitioners, dentists, physiotherapists, and other therapists and private practicing specialists. The database contains information on all health services provided in Denmark that were covered by National Health Insurance since January 1, 1990, as well as information on civil registration number, age, sex, region of residence, health provider, health-service code, and health-service description (examinations, tests, and procedures), and more.

Other linkages

The Danish and PHARMO systems also can be linked to many other databases and cohorts. For Denmark they have been described by Sorensen.[24] Linkage of other datasets to the PHARMO database are described on the PHARMO Institute website, including the national registry of traffic accidents (www.rijkswaterstaat.nl/), the national registry of driver licenses (www.cbr.nl), the national registry of kidney transplants (www.renine.nl), the Rotterdam study (http://www.epib.nl/research/ergo.htm), local and regional laboratories with records of digital ECGs, and a database including food constituents. These food constituents are collected in an annual survey of a project that focus on cardiovascular disease (MORGEN).[25] Both the Danish data network and the PHARMO network facilitate patient contact, to request their cooperation in providing patient-reported outcomes[26] or DNA.[27,28]

Strengths

Linkage possibilities in the Netherlands and Denmark, and also in Sweden, Iceland, and Norway,[5] are great in number. In theory, the databases can be expanded to include the complete countries. They constitute powerful resources for collecting detailed information for pharmacoepidemiologic research. A major strength of the pharmacy-based medical record linkage systems is the quality of the drug exposure information. Misclassification is not introduced by failure to fill prescriptions, as these data indicate what drugs were actually dispensed. The PHARMO system has an additional advantage, in that information on completeness of exposure does not depend on reimbursement status.

A second strength of the pharmacy-based medical record linkage systems discussed here is that, even in the absence of unique patient identifiers, databases can be linked with a high sensitivity and specificity using semideterministic or probabilistic record linkage methods. These methods have been shown to be extremely powerful in linking health-care registries and databases, and can also

be used to identify duplicates in patient router files. In the Netherlands, linkages have been established using probabilistic record linkage methods, comparable to linkages achieved in the Nordic countries, where unique identifiers in health care are available. Linkage possibilities are not restricted to the health-care domain, for example it is possible to link drug exposure files to food inventories or driver's license databases.

Another key asset of pharmacy-based medical record linkage systems is that exposure and outcomes are recorded separately, by the relevant health professionals, thereby minimizing the potential for ascertainment bias. Record linkage is also much less expensive than collecting data *de novo*.

Through contact with patients, biosamples and patient-reported outcomes can be obtained and linked to these databases. Pharmacy-based medical record linkage systems also have the advantage of having well-defined denominators, and including everyone in a given geographic area, regardless of whether information is recorded in hospitals, in the general practitioner's office, in clinical laboratories, or elsewhere. Researchers can focus on information in these databases that is most reliable and complete. Through linkage to nationwide registries and use of vital statistics, it is possible to assess the representativeness of findings for a complete nation.

Pharmacy files include prescriptions not only from GPs, but also from medical specialists, who are responsible for about a quarter of all prescriptions and more than half of all prescriptions given to patients with complicated diseases, although the percentage might differ slightly by country.[29] Omitting these prescriptions could cause bias by over-representation of healthier patients. This does not occur in pharmacy-based systems.

Pharmacy-based systems also contain great detail regarding type of drug, dose, and duration, permitting assessment of adherence and compliance to different types of medication, in relation to dose or route of administration.[30–37] They also provide precise and detailed product information—major assets for pharmacoepidemiologic studies.

Weaknesses

Inevitably, the weaknesses of medical record linkage systems are related to their strengths. The different organizations maintaining the patient-level databases must safeguard the integrity of the record linkage systems through rules of governance and access limitations. These organizations have to address many administrative issues, including confidentiality, conflicts of interest of health professionals and researchers, and procedures to ensure data quality. It is possible for individual organizations and data providers to prevent linkage for political and commercial reasons. For example, the PHARMO record linkage network faced the potential termination of the Dutch national hospital database.[38] This database, originally set up in 1963 as a scientific database of hospital admissions in the Netherlands, was used for years to allocate national budgets to hospitals. In 2006, the allocation of these budgets changed, and a new national database was created for this purpose. At this point hospitals did not recognize the need to contribute to the older national hospital database. Although most national statistics in the Netherlands depended on this database, it took many years of discussion to preserve it.[38]

Constant political experimenting with health-care systems is a major threat to the existence of medical record linkage systems. There is always a potential for a sudden, unexpected termination of registries or content of registries. For example, the Danish regional drug dispensing databases depend on the reimbursement status of drugs. As costs of sedatives, hypnotics, oral contraceptives, and laxatives are not reimbursed, these drugs cannot be studied. Although these data systems are not claims databases, they are used to claim reimbursement of dispensings to the different health insurance companies that require full identification of patients, their enrollees. If non-reimbursed drugs are dispensed without full patient identification, data may become incomplete. The possibility than exists that in the near future many inexpensive generics will cease to be reimbursed, which might impact on the completeness of recording dispensing in pharmacy-based medical record linkage databases and their

use for safety research. This problem is not specific to record linkage systems, but also affects claims databases. In the Netherlands, for instance, internet pharmacies and central distribution systems of biologicals can affect the completeness of data available to record linkage systems. Networks have to adapt and collaborate with these new drug sources to ensure data completeness.

Although the linkage domains seem unlimited, the complexity of pharmacy-based medical record linkage systems extends to the diverse governance structures responsible for protecting the privacy of patients and health-care professionals and for addressing potential conflicts of interest between them. Linked study datasets need to comply with the governance schemes of different organizations. Privacy laws in the European Union (see also Chapter 35) are not clear about whether needed data transfer is allowed, even after anonymization. Therefore, most registries do not give permission to distribute individual, anonymized patient records, whether linked or unlinked. The Health Improvement Network (THIN) database and the General Practice Research Database (GPRD) in the UK are exceptions (see Chapter 15).

The steady growth in computerization of health administration, and the high costs involved, as well as financial and other incentives, has led health-care providers to increase requests for reimbursement to provide data. Such price increases will increase the costs of studies, and may reduce the number of studies performed.

It must also be stressed that expert knowledge is required to work with these complex databases. The processes of collecting data and the associated effect on data validity and completeness are complex in record linkage systems. Researchers need to be trained to understand the implications of all these processes for pharmacoepidemiologic research. Even in less complicated environments, these issues have led highly qualified researchers to present conflicting results based on the same datasets. Education should be expanded beyond technical skills to include the nature of the data, limitations of interpreting study results, governance and privacy issues, as well as managing conflicts of interest. As one example, a student working in the PHARMO network asked a psychiatric patient for a buccal swab to obtain genetic material. Although the project complied with privacy laws and regulations, the patient complained to pharmacy personnel that his data was being used for research to such a extent that several pharmacies threatened to stop delivering data, jeopardizing the continuity of data collection and delaying other research. To address the need for expert management, the PHARMO Institute has a professional research staff that works with academic investigators in collecting and analyzing data. Academic investigators who regularly work with PHARMO data can sign a special agreement, allowing analysis of anonymous patient-based data outside the PHARMO Institute if they possess the required skills.

Managing the governance structures and statistical and content expertise needed to handle complex databases is time consuming and therefore expensive, despite the relatively low cost of data acquisition. These high operational costs constantly challenge the database holders. Increasing the size of study populations must be balanced against obtaining more in-depth information about each subject by establishing more linkages.

Record linkage systems potentially can cover 35 million European inhabitants in Northern Europe and the Netherlands, still far below the size of pharmacoepidemiologic databases in the US. Obviously, the primary advantage of the European systems is not size but detail.

Particular applications

Studies examining aspects of drug exposure

One study conducted in the PHARMO system showed an increased persistence rate (which might be expected to prolong treatment effect) for weekly versus daily alendronate (relative risk: 1.84, 95% CI: 1.65–2.20).[32] Dose–response relationships can also be examined. For example, dose–response relationships have been studied for statins and oral antidiabetic drugs concerning goal attainment of low density lipoprotein cholesterol (LDL-c) and

HbA1c levels, respectively.[30,31,39] A study performed by Eussen *et al*. linked medication use to responses to a food questionnaire. They studied the relationship between the intake of phytosterol/ phytostanol-enriched margarines in relation to persistence with statins, and found that overall statin discontinuation rates were not significantly different between users and non-users of enriched margarine. However, in the subgroup of starters, combination users had a higher risk of discontinuing statin therapy than single-component product users within 12 months (adjusted hazard ratio [HR]: 2.52, 95% CI: 1.06–6.00).[25] Duration–response relationships can also be studied. For example, Erichsen studied the relationship between statin use and gallstone formation, and found that among current users, the adjusted odds ratios associating statin use with the occurrence of gallstone disease were 1.17 (95% CI: 1.06–1.30) for those who had one to four prescriptions, 0.89 (95% CI: 0.80–0.97) for those who had five to nineteen prescriptions, and 0.76 (95% CI: 0.69–0.84) for those who had 20 or more total prescriptions.[40] In former users, the corresponding adjusted odds ratios were 1.24 (95% CI: 1.11–1.39), 0.97 (95% CI: 0.86–1.10), and 0.79 (95% CI: 0.64–0.97), respectively. The use of other lipid-lowering drugs did not show similar associations.

Studies using clinical laboratory files

Results obtained from GP PHARMO medical records showed that GPs tend to over-record abnormal values (as they obviously are of interest for further diagnostic and treatment decisions), compared to normal values. Average total cholesterol levels recorded in the GP data were 7.1 mmol/L, while the average total cholesterol values for the same patients were 5.6 mmol/L in the clinical laboratory files. Only 70% of the total cholesterol tests could be found in the GP files, measured against the clinical laboratory files. The differences were explained by tests ordered by medical specialists and by under-reporting of tests for patients who had reached their treatment goal (unpublished data).

Linking clinical laboratory data to drug exposure data from community pharmacies permits detailed research into the effect of drugs on biochemical parameters. For example, a study of patients with type 2 diabetes starting insulin compared glycemic control between those initiated on insulin detemir and those initiated on insulin glargine.[41] One year after start of insulin, there was no difference in mean HbA1c level or proportion of patients at goal (HbA1c < 7%) between users of the two preparations. Kornum *et al*. studied whether diabetes is a risk factor for hospitalization with pneumonia and assessed the impact of HgbA1c level on such risk. The adjusted relative risk for pneumonia-related hospitalization among subjects with diabetes was 1.26 (95% CI: 1.21–1.31) compared with non-diabetic individuals. The adjusted relative risk was 4.43 (95% CI: 3.40–5.77) for subjects with type 1 diabetes and 1.23 (95% CI: 1.19–1.28) for subjects with type 2 diabetes. Compared with subjects without diabetes, the adjusted relative risk was 1.22 (95% CI: 1.14–1.30) for diabetic subjects whose HgbA1c level was less than 7% and 1.60 (95% CI: 1.44–1.76) for diabetic subjects whose HbA1c level was greater than 9%. They concluded that poor long-term glycemic control among patients with diabetes increases the risk of hospitalization with pneumonia.[42] The impact of poor adherence on HgbA1c goal attainment was also studied using data from the PHARMO database. In patients starting oral glucose-lowering drugs (OGLD), the effect of non-persistent OGLD use on HbA1c goal attainment (<7%) was quantified, revealing that non-persistent patients were about 20% less likely to attain the goal compared to persistent patients.[30]

Lipid levels in statin users also have been studied using pharmacy-based databases. Heintjes *et al*. studied LDL-c goal attainment among daily users of different statins.[39] The proportion of patients attaining cholesterol goals was 75% for rosuvastatin, 68% for atorvastatin, 56% for simvastatin, and 42% for pravastatin. Dose comparisons showed greater LDL-c reduction and increased goal attainment for rosuvastatin 10 mg compared to other statins at most doses (adjusted $p < 0.05$). These results show clear dose–response relationships between different statins given in different doses, and extend reported clinical trial results to a real-world setting.[39] Use of serum cholesterol

measurements and LDLs among patients with a history of myocardial infarction (MI) in a Danish population was investigated using the OPED database. The investigators found insufficient use of lipid-lowering drug treatment in patients with established coronary heart disease.[43,44] Studies are currently underway to assess the relationship between statin use, LDL-c and high-density lipoprotein cholesterol (HDL-c) levels, and hospitalizations for cardiovascular conditions. Another study is assessing the relationship between risk factors such as HbA1c, creatinine, c-reactive protein, and cardiovascular complications of diabetes.

Pathology

The linkage of PHARMO and PALGA (the Dutch nationwide registry of histo- and cytopathology)[22,45] and counterpart databases in Denmark allow the study of the relationship between drug exposure and morbidity, assessed by pathology specimens retrieved via biopsy or resection. For example, in a study investigating estrogen exposure, retrieved via the PHARMO community pharmacy database, and the outcome of melanoma, retrieved via the PALGA database, a relationship between risk of cutaneous melanoma and cumulative dose of estrogens was identified.[46] In addition to the conclusion that exposure to certain drugs might increase the risk of cancer, drugs have also been found that prevent, reverse, suppress, or delay premalignant lesions. Recent studies have shown that both statins[47] and non-steroidal anti-inflammatory drugs (NSAIDs)[48] are associated with reduced incidence and progression of melanoma. The pathology registry in the Netherlands has a direct link with laboratories, including specimens stored locally in paraffin. More than 70% of laboratories have given researchers access for typing DNA, in order to study whether statin use in SMAD4-gene carriers protects against colorectal cancer (CRC). The study is still ongoing. The same linkage modality in Denmark has been used to investigate the relationship between proton pump inhibitor (PPI) use and CRC. After comparing the most intense users of PPI (more than every other day) to never or rare users, investigators found that neither short-term user (adjusted odds ratio [OR] 1.07, 95% CI: 0.86–1.34) nor long-term user (>7 years; adjusted OR 1.09, 95% CI: 0.58–2.06) was associated with an increased risk of cancer. They concluded that the use of PPIs in clinical practice does not measurably increase CRC risk.

Finally, linkage of pharmacy databases to pathology registries can be used to study patients' drug use after a cancer diagnosis. An observational study measured the association between tamoxifen adherence and recurrence of breast cancer,[49] examining whether concomitant use of CYP2D6 inhibitors, such as selective serotonin reuptake inhibitors, together with low tamoxifen adherence, may negatively impact tamoxifen efficacy in breast cancer patients. No association between concomitant CYP2D6 inhibitor use and breast cancer recurrence was observed (adjusted HR: 0.87, 95% CI: 0.42–1.79). Poor tamoxifen adherence was associated with shorter event-free time (adjusted HR: 0.987, 95% CI: 0.975–0.999). The study did not show an association between concomitant CYP2D6 inhibitor use and breast cancer recurrence among patients treated with adjuvant tamoxifen, despite the strong biologic rationale. This study provides evidence, for the first time, that poor tamoxifen adherence is associated with increased risk of breast cancer events.

Cancer

Relationships between drug use and cancer, either occurrence or treatment, also can be studied using linked data from hospitals, GP files, and clinical pathology laboratory findings. Several studies have examined the effectiveness of cytostatics and comedication in cancer patients. The added value of linkage to the cancer registries, in addition to hospital registries, is that incident cases can be identified and extra information is available, such as cancer type, staging of the cancer, and non-drug treatment. This information is of great importance in studying the relationship between drug use and cancer, which may have a long lag-time.

The possible relationship between the use of oral glucocorticoids and increased risk of basal cell carcinoma (BCC), squamous cell carcinoma (SCC), malignant melanoma (MM), and non-Hodgkin's lymphoma (NHL) was studied using the AUPD database. The results showed slightly elevated risk

estimates for BCC: incidence rate ratio (IRR): 1.15 (95% CI: 1.07–1.25), SCC IRR: 1.14 (95% CI: 0.94–1.39), MM IRR: 1.15 (95% CI: 0.94–1.41), and NHL IRR: 1.11 (95% CI: 0.85–1.46). These results support an overall association between glucocorticoid use and the risk of BCC.[50]

As another example, Sukel et al. studied the incidence of cardiovascular events among breast cancer patients after chemotherapy. After 1 year of follow-up, the incidence rate of cardiovascular events was 69/1000 person-years for patients given cardiotoxic chemotherapy and 98/1000 person-years for patients with non-cardiotoxic chemotherapy. These rates were not statistically significantly different (HR: 0.74, 95% CI: 0.39–1.41). This study also showed similar cardiovascular incidence rates during follow-up for breast cancer patients treated with cardiotoxic and non-cardiotoxic chemotherapy. It appeared that specialists took pre-existing cardiovascular diseases into account in choosing treatment regimens.[51] Recent studies of treatment patterns in breast cancer patients reported the type of chemotherapy regimens administered to early-stage and metastatic breast cancer patients (van Herk-Sukel MPP, van de Poll-Franse LV, Creemers GJ, Lemmens VEPP, van der Linden PD, Herings RMC, et al. Major changes in type of chemotherapy regimens administered to patients with early or metastatic breast cancer during 2000–2008 in the Southeastern Netherlands. Unpublished data.) and showed continuation of endocrine treatment after diagnosis with early stage breast cancer.[23] Ewertz et al. studied the risk of developing breast cancer in relation to the use of hormonal replacement therapy (HRT), reporting an elevated risk (1.61, [95% CI: 1.38–1.88]) for HRT use in women older than 50 years.[52] The risk increased with increasing duration of use and decreased with time since last HRT prescription, with the risk ratio reaching unity after 5 years.

In-hospital drug use

In addition to cancer treatment, treatment of other morbidities during hospitalization can be studied through linkage of the hospital registry data to inpatient pharmacy data. For example, opioid use has been examined in patients admitted for genitourinary, digestive, or abdominal surgery.[53] Use of (nico)morphine was associated with the risk for developing postoperative paralytic ileus (POI) (OR: 12.1, 95% CI: 5.4–27.1). The association between opioids and POI was most obvious in patients with abdominal surgery (OR: 33.8, 95% CI: 6.2–184.6) and patients without colon/ colorectal/ rectal tumors (OR: 13.2, 95% CI: 5.7–30.3). A clear association was found between the use of opioids and the risk for POI, as coded in the Dutch hospital registry. In another study, the relationship between initial antibiotic treatment of secondary intra-abdominal infections and related outcomes was assessed. It was found that inappropriate initial antibiotic treatment was associated with a 3.4-fold (95% CI: 1.3–9.1) risk of clinical failure. The length of hospital stay and costs of hospitalization were significantly increased for patients with antibiotic failure.[54]

Birth registries

The effect of drugs and adverse drug reactions in children and the effect of drugs during pregnancy are an important area of study, as these relationships can only be examined to a very limited extent, if at all, in clinical trials. Linkage to perinatal registries permits detailed research into drug exposures and co-morbidities during pregnancy and short- and long-term health status of offspring, topics that are well established in the Northern countries.[55–65] In their study of the safety of metoclopramide use during pregnancy, Sorensen *et al*. found no major differences in the risk of malformations (OR: 1.11, 95% CI: 0.6–2.1); low birth weight (OR: 1.79, 95% CI: 0.8–3.9), or preterm delivery (OR: 1.02, 95% CI: 0.6–1.7).[62] Olesen *et al*. studied the risk of sumatriptan use during pregnancy and found that the risk of preterm delivery was elevated among women exposed to sumatriptan compared with migraine controls (OR: 6.3, 95% CI: 1.2–32.0) and healthy women (OR: 3.3, 95% CI: 1.3–8.5). The odds ratio for having a newborn with low birth weight was increased (OR: 3.0, 95% CI: 1.3–7.0) for all migraine patients who delivered at term compared with healthy women's pregnancies.[61] These findings may be due to drug exposure, but they may also reflect the impact of the disease itself

rather than treatment. Confounding or chance could also underlie the findings. These examples illustrate the complexity of controlling bias and confounding, and the difficulty of achieving adequate statistical power because of low numbers of exposed pregnancies. Collaboration among different databases is needed to obtain enough exposed births. The Dutch perinatal registry was linked in 2010 to the PHARMO registries, providing access to more than 175 000 births (with follow-up >10 years) and pregnancies. Similar studies will be possible in Denmark. Currently, the linkage has been validated, and a pilot study comparing preterm and full-term infants revealed that, in the first year of life, preterm infants are up to two times as likely as full-term infants to be hospitalized or use medication, especially related to respiratory diseases.[56]

The future

The major future challenge of medical record linkage systems is to protect the structures that make possible linkage between different databases. The Northern countries and the Netherlands, together covering almost 35 million inhabitants, have the potential to build extremely detailed databases for use in pharmacoepidemiologic research. The databases described above already cooperate in several projects financed by the European Commission (FP7) (http://cordis.europa.eu/fp7/) and are part of the European Network for Centres of Pharmacoepidemiology and Pharmacovigilance (ENCePP) (www.encepp.eu). However, given the complexity of EU privacy protection laws, governance models are needed to safeguard access to medical record linkage systems. Data sharing and integration of different national resources will be a true challenge in coming years. Experience in the Northern countries and the Netherlands show that independent organizations with stringent control of access can help to safeguard medical record linkage systems.

Future goals are to add patient-reported outcomes and DNA to the pharmacy-based medical record linkage systems. Studies in the Netherlands have shown the feasibility of these linkages and their capacity to support pharmacogenetic research (see Chapter 34), outcomes research, research on quality of life (see Chapter 39), as well as research on adverse outcomes and risk factors that will contribute to a better understanding of the risk and effectiveness of drugs in daily life.

References

1. Gaist D, Sorensen HT, Hallas J. The Danish prescription registries. *Dan Med Bull* 1997; **44**: 445–8.
2. Furu K. Establishment of the nationwide Norwegian Prescription Database (NorPD)—new opportunities for research in pharmacoepidemiology in Norway. *Norsk Epidemiologi* 2008; **18**: 129–36.
3. Klaukka T. The Finnish database on drug utilisation. *Norwegian J Epidemiol* 2001; **11**: 19–22.
4. Wettermark B, Hammar N, Fored CM, Leimanis A, Otterblad Olausson P, Bergman U, *et al*. The new Swedish Prescribed Drug Register—opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf* 2007; **16**: 726–35.
5. 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**: 86–94.
6. Morris AD, Boyle DI, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, *et al*. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. *BMJ* 1997; **315**: 524–8.
7. Herings RMC. *PHARMO: A Record Linkage System for Postmarketing Surveillance of Prescription Drugs in the Netherlands* [Doctorate]. Utrecht: Utrecht University, 1993.
8. Klinisk Epidemiologisk Afdeling. www.kea.au.dk/ (accessed June 2011).
9. PHARMO. www.pharmo.com (accessed June 2011).
10. Hallas J. Conducting pharmacoepidemiologic research in Denmark. *Pharmacoepidemiol Drug Saf* 2001; **10**: 619–23.
11. Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. *Clin Epidemiol* 2010; **2**: 273–9.
12. Wagner G, Newcombe HB. Record linkage. Its methodology and application in medical data processing. *Methods Inf Med* 1970; **9**: 121–38.

13. DuVall SL, Kerber RA, Thomas A. Extending the Fellegi-Sunter probabilistic record linkage method for approximate field comparators. *J Biomed Inform* 2010; **43**: 24–30.
14. Herzog TN, Scheuren F, Winkler WE. *Data Quality and Record Linkage Techniques*. New York: Springer, 2007, p. 227.
15. Strom B, Kimmel S. *Textbook of Pharmacoepidemiology*, 4th edn. West Sussex, England: John Wiley & Sons, 2006.
16. Mason CA, Tu S. Data linkage using probabilistic decision rules: a primer. *Birth Defects Res A Clin Mol Teratol* 2008; **82**: 812–21.
17. Goehring R. Identification of patients in medical data bases—soundex codes versus match code. *Med Inform* (Lond) 1985; **10**: 27–34.
18. Baldwin JA. The Oxford Record Linkage Study as a medical information system. *Proc R Soc Med* 1972; **65**: 237–9.
19. Tromp M, Ravelli AC, Meray N, Reitsma JB, Bonsel GJ. An efficient validation method of probabilistic record linkage including readmissions and twins. *Methods Inf Med* 2008; **47**: 356–63.
20. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol* 2010; **64**: 565–72.
21. van Herk-Sukel MP, van de Poll-Franse LV, Lemmens VE, Vreugdenhil G, Pruijt JF, Coebergh JW, *et al*. New opportunities for drug outcomes research in cancer patients: the linkage of the Eindhoven Cancer Registry and the PHARMO Record Linkage System. *Eur J Cancer* 2010; **46**: 395–404.
22. Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, *et al*. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol* 2007; **29**: 19–24.
23. van Herk-Sukel MP, van de Poll-Franse LV, Voogd AC, Nieuwenhuijzen GA, Coebergh JW, Herings RM. Half of breast cancer patients discontinue tamoxifen and any endocrine treatment before the end of the recommended treatment period of 5 years: a population-based analysis. *Breast Cancer Res Treat* 2010; **122**: 843–51.
24. Sorensen HT, Christensen T, Kjeldahl Schlosser H, Pedersen L. *Use of Medical Databases in Clinical Epidemiology*. Aarhus University, Denmark, 2008.
25. Eussen SR, de Jong N, Rompelberg CJ, Garssen J, Verschuren WM, Klungel OH. Effects of the use of phytosterol/-stanol-enriched margarines on adherence to statin therapy. *Pharmacoepidemiol Drug Saf* 2010; **19**: 1225–32.
26. Penning-van Beest FJ, van den Haak P, Klok RM, Prevoo YF, van der Peet DL, Herings RM. Quality of life in relation to constipation among opioid users. *J Med Econ* 2010; **13**: 129–35.
27. Peters BJ, Rodin AS, Klungel OH, Stricker BH, de Boer A, Maitland-van der Zee AH. Variants of ADAMTS1 modify the effectiveness of statins in reducing the risk of myocardial infarction. *Pharmacogenet Genomics* 2010; **20**: 766–74.
28. Peters BJ, Rodin AS, Klungel OH, van Duijn CM, Stricker BH, van't Slot R, *et al*. Pharmacogenetic interactions between ABCB1 and SLCO1B1 tagging SNPs and the effectiveness of statins in the prevention of myocardial infarction. *Pharmacogenomics* 2010; **11**: 1065–76.
29. Herings RMC. Medicatiebewaking gebaat bij samenwerking. *Pharm Weekblad* 1997; **132**: 1030–3.
30. Penning-van Beest FJA, van der Bij S, Erkens JA, Kessabi S, Groot M, Herings RM. Effect on non-persistent use of oral glucose-lowering drugs on HbA1c goal attainment. *Curr Med Res Opin* 2008; **24**: 2523–9.
31. Penning-van Beest FJ, Termorshuizen F, Goettsch WG, Klungel OH, Kastelein JJ, Herings RM. Adherence to evidence-based statin guidelines reduces the risk of hospitalizations for acute myocardial infarction by 40%: a cohort study. *Eur Heart J* 2006; **28**: 154–9.
32. Penning-van Beest FJ, Goettsch WG, Erkens JA, Herings RM. Determinants of persistence with bisphosphonates: a study in women with postmenopausal osteoporosis. *Clin Ther* 2006; **28**: 236–42.
33. Breekveldt-Postma NS, Herings RM. Persistence with antihypertensives related to formulation: the case of nifedipine. *Ann Pharmacother* 2005; **39**: 237–42.
34. Hansen DG, Vach W, Rosholm JU, Sondergaard J, Gram LF, Kragstrup J. Early discontinuation of antidepressants in general practice: association with patient and prescriber characteristics. *Fam Pract* 2004; **21**: 623–9.
35. Herings RM, Erkens JA. Increased suicide attempt rate among patients interrupting use of atypical antipsychotics. *Pharmacoepidemiol Drug Saf* 2003; **12**: 423–4.
36. Larsen J, Andersen M, Kragstrup J, Gram LF. High persistence of statin use in a Danish population: compliance study 1993–1998. *Br J Clin Pharmacol* 2002; **53**: 375–8.

37. 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**: 463–71.
38. Stricker BH, Herings RM. Pleidooi voor het behoud van de Landelijke Medische Registratie voor betrouwbare informatie over de volksgezondheid en de gezondheidszorg. *Ned Tijdschr Geneeskd* 2006; **150**: 1916–7.
39. Heintjes EM, Hirsch MW, Van der Linden MW, O'Donnell JC, Stalenhoef AF, Herings RMC. LDL-C reductions and goal attainment among naive statin users in the Netherlands: real life results. *Curr Med Res Opin* 2008; **24**: 2241–50.
40. Erichsen R, Froslev T, Lash TL, Pedersen L, Sorensen HT. Long-term statin use and the risk of gallstone disease: A population-based case–control study. *Am J Epidemiol* 2011; **173**: 162–70.
41. Heintjes EM, Thomsen TL, Penning-van Beest FJ, Christensen TE, Herings RM. Glycemic control and long-acting insulin analog utilization in patients with type 2 diabetes. *Adv Ther* 2010; **27**: 211–22.
42. Kornum JB, Thomsen RW, Riis A, Lervang HH, Schonheyder HC, Sorensen HT. Diabetes, glycemic control, and risk of hospitalization with pneumonia: a population-based case–control study. *Diabetes Care* 2008; **31**: 1541–5.
43. Kanstrup H, Lassen JF, Heickendorff L, Lauritzen T, Larsen ML. Quality of lipid-lowering therapy in patients with ischaemic heart disease: a register-based study in 3477 patients. *J Intern Med* 2004; **255**: 367–72.
44. Larsen J, Andersen M, Bjerrum L, Kragstrup J, Gram LF. Insufficient use of lipid-lowering drugs and measurement of serum cholesterol among patients with a history of myocardial infarction. *J Cardiovasc Risk* 2003; **10**: 61–4.
45. PALGA. *The nationwide network and registry of histo- and cytopathology in the Netherlands*. www.palga.nl (accessed June 2011).
46. Koomen ER, Joosse A, Herings RM, Casparie MK, Guchelaar HJ, Nijsten T. Estrogens, oral contraceptives and hormonal replacement therapy increase the incidence of cutaneous melanoma: a population-based case–control study. *Ann Oncol* 2009; **20**: 358–64.
47. Koomen ER, Joosse A, Herings RM, Casparie MK, Bergman W, Nijsten T, *et al*. Is statin use associated with a reduced incidence, a reduced Breslow thickness or delayed metastasis of melanoma of the skin? *Eur J Cancer* 2007; **43**: 2580–9.
48. Joosse A, Koomen ER, Casparie MK, Herings RM, Guchelaar HJ, Nijsten T. Non-steroidal anti-inflammatory drugs and melanoma risk: large Dutch population-based case-control study. *J Invest Dermatol* 2009; **129**: 2620–7.
49. Dezentje VO, van Blijderveen NJ, Gelderblom H, Putter H, van Herk-Sukel MP, Casparie MK, *et al*. Effect of concomitant CYP2D6 inhibitor use and tamoxifen adherence on breast cancer recurrence in early-stage breast cancer. *J Clin Oncol* 2010; **28**: 2423–9.
50. Jensen AO, Thomsen HF, Engebjerg MC, Olesen AB, Friis S, Karagas MR, *et al*. Use of oral glucocorticoids and risk of skin cancer and non-Hodgkin's lymphoma: a population-based case-control study. *Br J Cancer* 2009; **100**: 200–5.
51. Sukel MP, Breekveldt-Postma NS, Erkens JA, van der Linden PD, Beiderbeck AB, Coebergh JW, *et al*. Incidence of cardiovascular events in breast cancer patients receiving chemotherapy in clinical practice. *Pharmacoepidemiol Drug Saf* 2008; **17**: 125–34.
52. Ewertz M, Mellemkjaer L, Poulsen AH, Friis S, Sorensen HT, Pedersen L, *et al*. Hormone use for menopausal symptoms and risk of breast cancer. A Danish cohort study. *Br J Cancer* 2005; **92**: 1293–7.
53. Goettsch WG, Sukel MP, van der Peet DL, van Riemsdijk MM, Herings RM. In-hospital use of opioids increases rate of coded postoperative paralytic ileus. *Pharmacoepidemiol Drug Saf* 2007; **16**: 668–74.
54. Sturkenboom MC, Goettsch WG, Picelli G, in 't Veld B, Yin DD, de Jong RB, *et al*. Inappropriate initial treatment of secondary intra-abdominal infections leads to increased risk of clinical failure and costs. *Br J Clin Pharmacol* 2005; **60**: 438–43.
55. Ratanajamit C, Vinther Skriver M, Jepsen P, Chongsuvivatwong V, Olsen J, Sorensen HT. Adverse pregnancy outcome in women exposed to acyclovir during pregnancy: a population-based observational study. *Scand J Infect Dis* 2003; **35**: 255–9.
56. Olesen C, Sondergaard C, Thrane N, Nielsen GL, de Jong-van den Berg L, Olsen J. Do pregnant women report use of dispensed medications? *Epidemiology* 2001; **12**: 497–501.
57. Olesen C, Thrane N, Nielsen GL, Sorensen HT, Olsen J. A population-based prescription study of asthma drugs during pregnancy: changing the intensity of asthma therapy and perinatal outcomes. *Respiration* 2001; **68**: 256–61.

58. Fonager K, Larsen H, Pedersen L, Sorensen HT. Birth outcomes in women exposed to anticonvulsant drugs. *Acta Neurol Scand* 2000; **101**: 289–94.
59. Sorensen HT, Johnsen SP, Larsen H, Pedersen L, Nielsen GL, Moller M. Birth outcomes in pregnant women treated with low-molecular-weight heparin. *Acta Obstet Gynecol Scand* 2000; **79**: 655–9.
60. Larsen H, Nielsen GL, Sorensen HT, Moller M, Olsen J, Schonheyder HC. A follow-up study of birth outcome in users of pivampicillin during pregnancy. *Acta Obstet Gynecol Scand* 2000; **79**: 379–83.
61. Olesen C, Steffensen FH, Sorensen HT, Nielsen GL, Olsen J. Pregnancy outcome following prescription for sumatriptan. *Headache* 2000; **40**: 20–4.
62. Sorensen HT, Nielsen GL, Christensen K, Tage-Jensen U, Ekbom A, Baron J. Birth outcome following maternal use of metoclopramide. The Euromap study group. *Br J Clin Pharmacol* 2000; **49**: 264–8.
63. Nielsen GL, Sorensen HT, Thulstrup AM, Tage-Jensen U, Olesen C, Ekbom A. The safety of proton pump inhibitors in pregnancy. *Aliment Pharmacol Ther* 1999; **13**: 1085–9.
64. Sorensen HT, Nielsen GL, Olesen C, Larsen H, Steffensen FH, Schonheyder HC, *et al*. Risk of malformations and other outcomes in children exposed to fluconazole in utero. *Br J Clin Pharmacol* 1999; **48**: 234–8.
65. Olesen C, Steffensen FH, Nielsen GL, de Jong-van den Berg L, Olsen J, Sorensen HT. Drug use in first pregnancy and lactation: a population-based survey among Danish women. The EUROMAP group. *Eur J Clin Pharmacol* 1999; **55**: 139–44.

CHAPTER 19
Case–Control Surveillance

Lynn Rosenberg, Patricia F. Coogan, and Julie R. Palmer
Slone Epidemiology Center at Boston University, Boston, MA, USA

Introduction

Premarketing trials for safety and efficacy are too small to detect any but common adverse effects and too brief to detect effects that occur long after use.[1] For cancers especially, there is a need to identify unintended effects that occur after long latent intervals or durations of use. Follow-up strategies that link pharmacy data with outcome data have been useful for monitoring the unintended health effects of prescription medications, but information on preparations sold over-the-counter has been limited.[2] Drugs previously available only by prescription (e.g., omeprazole) are increasingly available on a non-prescription basis.[3,4] In addition, dietary supplements (e.g., herbals) are used widely;[5] these too are sold over-the-counter and they do not have to be shown to be safe before being marketed.[6–15] Thus, there is a need to monitor the unintended health effects of non-prescription medications and dietary supplements as well as those of prescription drugs. Case–control surveillance (CCS), a surveillance system based on case–control methods, has these capabilities.

Description

Overview

CCS was begun in 1976 with funding from the US Food and Drug Administration, with the aim of assessing medication use in relation to both non-malignant and malignant illnesses. The focus was subsequently changed to cancer, with funding provided by the National Cancer Institute from 1988 to 2009. Patients with recently diagnosed cancer or non-malignant conditions were interviewed in a set of participating hospitals for a lifetime history of regular medication use and information on potential confounding or modifying factors. To assess associations of medication use with diseases of interest, cases with the selected diseases were compared to appropriate controls with other illnesses in case–control analyses. The large CCS database, described below, has been used for the assessment of associations of many medications and risk factors with risk of many illnesses. It is available for further analyses.

Case and control accrual and classification

Data collection took place at collaborating hospitals in several geographic areas from 1976 to 2009. Institutional review board approval was obtained from collaborating institutions and the study complied with Health Insurance Portability and Accountability Act (HIPAA) requirements. Because the vast majority of patients were accrued in hospitals in Boston, Baltimore, New York, and Philadelphia, the database has been restricted to patients from these areas. After obtaining written informed consent, nurse-interviewers trained and

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

employed by CCS interviewed adult patients, 21–79 years of age, admitted for recently diagnosed cancers or non-malignant disorders; patients with non-malignant illnesses serve as a pool of potential controls in the case–control analyses and may themselves be of interest as outcomes (e.g., cholecystitis,[16] pelvic inflammatory disease[17]). Because patients with conditions of acute onset (e.g., traumatic injury, appendicitis) are suitable controls in many analyses, they were selectively accrued. For more chronic conditions (e.g., orthopedic disorders, kidney stones), recruitment was confined to patients whose diagnosis was made within the previous year. To guard against referral bias, only patients living in areas within approximately 50 miles of the hospital were eligible for inclusion. If several patients were available for interview, patients with diseases of special interest were selectively interviewed according to a priority list. The interview setting—a hospital or clinic room—was the same for cases and controls. Many diseases and exposures were assessed and cases in one analysis might be controls in another. Thus, the interviewers were unaware of the "case" or "control" status or the exposure status of the patient interviewed. A total of 66 445 patients were interviewed. The participation rate exceeded 85%.

A copy of the discharge summary was obtained for all interviewed patients and the pathology report for patients with cancer. These were reviewed and abstracted in the central office by the study nurse-coordinator, blind to exposure category.

Table 19.1 gives the numbers at various cancer sites among the 26 823 interviewed patients with recently diagnosed primary cancers. There are 8002 patients with breast cancer, 3124 with large bowel cancer, 2001 with lung cancer, at least 1000 each with malignant melanoma, ovarian cancer, or endometrial cancer, and at least 500 each with leukemia, bladder cancer, pancreatic cancer, non-Hodgkin's lymphoma, renal cancer, or bone/ connective tissue cancer.

Table 19.2 lists common diagnostic categories among the 39 622 patients admitted for non-malignant conditions. Patients with non-malignant diagnoses serve as a pool of controls for analyses of various cancers, and the diagnoses can themselves be assessed as the outcome of interest. Among the most common non-malignant diagnoses are traumatic injury (e.g., fractured arm), benign neoplasms, acute infections (e.g., appendicitis), orthopedic disorders (e.g., disc disorder), gallbladder disease, and hernias.

Table 19.1 Patients with cancer interviewed in CCS

Cancer	Female (n)	Male (n)	Total (n)
Breast	7984	18	8002
Large bowel	1492	1632	3124
Lung	845	1156	2001
Prostate	—	1996	1996
Malignant melanoma	897	747	1644
Ovary	1167	—	1167
Endometrium	1079	—	1079
Leukemia	452	478	930
Bladder	190	604	794
Kidney/ kidney pelvis	230	462	692
Pancreas	313	373	686
Non-Hodgkin's lymphoma	283	302	585
Bone/connective tissue	302	232	534
Testis	—	477	477
Stomach	155	178	333
Hodgkin's disease	171	149	320
Esophagus	74	239	313
Gallbladder	79	80	159
Vulva	151	—	151
Liver	40	31	71

Drug information and classification

It is not feasible to ask specifically about thousands of individual medications. Patients in CCS were

Table 19.2 Patients with non-malignant conditions interviewed in CCS

Non-malignant condition	Number
Fracture	2986
Other injury	2749
Uterine fibroid	1818
Benign neoplasm	1625
Cholecystitis	1544
Displacement of intervertebral disc	1434
Ovarian cyst	1418
Hernia	1135
Appendicitis	977
Cholelithiasis	943
Calculus of kidney and ureter	756
Pelvic inflammatory disease	711
Benign prostatic hypertrophy	640
Ectopic pregnancy	470
Diverticulitis	420
Endometriosis	390
Cellulitis	380
Pancreatitis	290
Spinal stenosis	250
Bowel obstruction	250

questioned about regular use for 43 indication or drug categories, for example headache, cholesterol-lowering, oral contraception, menopausal symptoms, herbals/ dietary supplements. For each episode of use, the preparation name and the timing, duration, and frequency of use were recorded. The drug dose was recorded only when it was part of the brand name; for example for oral contraceptives and conjugated estrogens, the brand name may indicate the dosage.

The Slone Drug Dictionary (www.bu.edu/slone/) maintained by our research group was used to code all medications and dietary supplements reported by CCS participants. The dictionary is a computerized linkage system composed of individual medicinal agents and multicomponent products. Each is assigned a specific code number. The codes are linked to the active ingredients, which permits the investigation of medication exposures according to the component (e.g., acetaminophen), the drug class (e.g., analgesic/antipyretic), whether the product is a single or multiple entity product, and the specific product. The Slone Drug Dictionary contains over 14 500 single-entity medicinal agents and more than 10 000 multicomponent products. Numerous coalitions have been formed (e.g. selective serotonin reuptake inhibitors, calcium channel blockers, tricyclic antidepressants, thiazide diuretics, benzodiazepines, and beta-adrenergic blockers).

Information on factors other than drugs

CCS collected information on many factors that may confound or modify drug–disease associations, including descriptive characteristics (e.g., age, height, current weight, weight 10 years ago, weight at age 20, years of education, marital status, racial/ethnic group), habits (cigarette smoking, alcohol consumption, coffee consumption), gynecologic and reproductive factors (age at first birth, parity, age at menarche and menopause, and type of menopause), medical history (cancer, hypertension, diabetes, other serious illnesses, vasectomy, hysterectomy, oophorectomy), family history of cancer, use of medical care (e.g., number of visits to a physician in each of the previous 2 years). These variables can also be assessed as risk factors in their own right.

Data analysis

Case and control specification

For each analysis, the case series is defined, for example women with invasive primary ovarian cancer diagnosed less than a year before admission and documented in the pathology report. The proper selection of controls in hospital-based studies is essential for validity. For the particular exposure of interest, the controls selected should have been admitted for conditions that are not caused, prevented, or treated by that exposure.[18–20] We select several appropriate diagnostic categories with sufficient numbers to allow for examination of the prevalence of the exposure of interest across the categories. If our judgment about control selection is correct, the prevalence of that exposure is uniform across the categories.

Aspects of drug use

The information collected in CCS on episodes of drug and supplement use (name and timing, frequency, and duration of use) allows for the assessment of many aspects of use. For example, for breast cancer, drug use at potentially vulnerable times during reproductive life (e.g., before the birth of the first child) can be assessed. The particular drug or drug regimen may also be relevant; for example the risk of endometrial cancer is increased by unopposed estrogen supplements, but little or not at all by combined use of estrogen with a progestogen.[21] The timing of use may be relevant; for example in accord with animal data, in our analysis of CCS data on non-steroidal anti-inflammatory drugs (NSAIDs) and large bowel cancer,[22] recent use was associated with a reduced odds ratio whereas past use was not. Latent intervals may or may not be relevant; for example our analyses of a non-drug exposure, vasectomy, in relation to the risk of 10 cancer sites found no trends according to the interval between vasectomy and the occurrence of the cancer.[23] It may be of interest to determine how long an increased or reduced odds ratio persists after an exposure has ended; for example CCS data indicated that an association of a reduced risk of ovarian cancer with oral contraceptive use persisted for 15–19 years after cessation of use.[24]

Some drugs, particularly non-prescription drugs such as acetaminophen, are often used sporadically. Sporadic use in the past probably cannot be reported accurately, and regular use may be more likely to play an etiologic role. Thus, we place our greatest reliance on regular use (e.g., at least four times a week for at least 3 months), and particularly on regular use for several years or more. The observation of stronger associations for more frequent or long duration medication use provides support for a causal role.

We have not collected information on the dose of drugs because of evidence of poor recall;[25] for example, women who have used several different brands of oral contraceptives often have difficulty remembering the brand and dosage.[26–30] However, the frequency and duration of use provide a useful measure of the intensity of exposure.

Odds ratio estimation with control of confounding factors

Odds ratios and 95% confidence intervals are estimated using multiple logistic regression analysis.[18] Potential confounding factors are controlled in the regression models if their inclusion materially alters the odds ratio, for example by 10% or more.

Effect modification

Effect modification is assessed by examining exposure–disease associations in subgroups and by statistical modeling, such as the use of interaction terms in logistic regression. For example, in our analysis of estrogen supplements in relation to risk of breast cancer, the overall findings were null but estrogen supplement use was associated with increased risk of breast cancer among thin women.[31]

Statistical power

Table 19.3 shows the sample sizes needed for 80% statistical power to detect a range of odds ratios for a range of exposure prevalences. CCS has excellent statistical power for the detection of associations that are of public health importance.

Discovery of unsuspected associations

Animal data have led to the search for and identification of new associations in CCS data. For example, experiments in rodents treated with car-

Table 19.3 Number of cases needed for detection of odds ratios of 1.5 to 4 with 80% or greater power for exposure prevalences of 0.25% to 15%; alpha = 0.05 (two tailed), control-to-case ratio = 4:1

Exposure prevalence in controls (%)	Odds ratio			
	1.5	2	3	4
15	380	115	40	25
10	520	150	50	30
5	950	270	85	45
3	1520	425	130	70
2	2235	620	185	100
1	4395	1205	360	185
0.5	8710	2385	700	360
0.25	17340	4740	1390	710

cinogens suggested that NSAIDs might reduce the occurrence of large bowel cancer. In an analysis of CCS data, risk of large bowel cancer was inversely associated with aspirin use,[22] an association that was confirmed in many subsequent studies.[32–36]

Associations have also been identified by systematic "screening" of the data, in which the prevalence of use of a particular drug, drug class, or other factor among patients with a particular illness is compared with the prevalence among patients with other illnesses. Examples of unexpected drug–disease associations from screening are oral contraceptive use with choriocarcinoma[37] and Crohn's disease;[38] these associations have received independent confirmation.[39,40] Screening of non-drug factors revealed an unexpected association between alcohol use and breast cancer,[41] confirmed in subsequent studies.[42–44] Associations that arise in the course of multiple comparisons may of course be due to chance. Even if associations are not due to chance, the magnitude of the association will tend to "regress to the mean" in subsequent studies. For these reasons, new associations need to be presented with caution.

Strengths

Capacity to assess non-prescription medications and supplements as well as prescription medications

CCS can assess the effects of prescription medications from multiple sources (e.g., family planning clinics, friends, and relatives) whereas monitoring systems that rely on pharmacy data can study only those medications that are prescribed or dispensed within the system. Another limitation of reliance on prescription data is that prescribed medications are sometimes not taken. CCS is the only surveillance system that can systematically assess use of non-prescription medications and dietary supplements. The prevalence of dietary supplement use has increased to a level where assessment of their potential effects on disease occurrence is of public health importance. Certain medications, such as oral contraceptives and non-contraceptive estrogens, have been well studied in relation to a range of conditions, but the effects of many other drugs have not. CCS has assessed associations with a wide range of medications, including angiotensin converting enzyme (ACE) inhibitors, acetaminophen, antidepressants, antihistamines, aspirin and other NSAIDs, benzodiazepines, beta-androgenic blockers, calcium channel blockers, diuretics, female hormone supplements, hydralazine, oral contraceptives, phenolphthalein-containing laxatives, phenothiazines, rauwolfia alkaloids, selective serotonin reuptake inhibitors, statins, thiazides, thyroid supplements, and vitamins/ minerals (see the list of publications in the Appendix).

Unintended effects of medications documented by CCS include adverse effects (e.g., increased risk of liver cancer[45] and breast cancer[46–48] associated with oral contraceptive use, and increased risk of localized and advanced endometrial cancer associated with postmenopausal estrogen supplement use[49]) and protective effects (e.g., reduced risks of ovarian[24] and endometrial cancer[50] associated with oral contraceptive use and reduced risks of colorectal cancer[22] and stomach cancer associated with aspirin use[51]). The safety of drugs has also often been documented by CCS after concerns had been raised about adverse effects. Experiments in rodents given phenolphthalein suggested increased risks of several cancers[52] but CCS findings suggested no increased risk.[53] A study suggested that calcium channel blockers increased the risk of several cancers[54,55] but results based on much larger numbers in CCS refuted that finding.[56] Animal data raised the concern that benzodiazepines increased the risk of several cancers, but results from CCS were null.[57–59] Animal data also raised the possibility of increased risks of cancer associated with hydralazine use, but again CCS results were null.[60,61] The CCS database remains available for further analyses of interest.

Assessment of effects after long intervals or durations of use

Effects of drugs may become evident only after many years. The case–control design used by CCS is useful for assessing the effects of exposures that occurred in the distant past or after long durations of exposure; for example CCS documented an

increased risk of endometrial cancer in association with use of estrogen supplements 15–19 years after cessation of use.[49] Cohort studies cannot make these assessments unless they have been in progress for many years.[18]

Control of confounding

Much information on drug–disease associations comes from non-randomized studies. Therefore, control of confounding is crucial for validity. Medication use is associated with many factors that are strongly associated with disease risk. CCS has information on important potential confounding factors that can be controlled in multivariable analyses. These factors include demographic characteristics, medical history, reproductive and gynecologic history, family history of cancer, tobacco and alcohol use, use of medical care, and use of other drugs.

Accurate outcome data

CCS collected the hospital discharge summary for all patients and the pathology report for patients with cancer, thus allowing for accurate classification of the diagnosis.

High statistical power

As shown in Table 19.3, CCS has high statistical power to assess the effects of exposures of public health importance. Small odds ratios associated with uncommon drug exposures can be detected for common cancers,[19] while moderate odds ratios associated with more common exposures can be detected for less common cancers. For very rare cancers, only relatively large effects can be detected for relatively common exposures.

Productivity and substantive findings

CCS has been highly productive, with 95 papers published on a variety of exposures and outcomes (see Appendix).

Weaknesses

Selection bias

Biased selection of cases and controls may occur because of referral bias in hospital-based studies such as CCS. Bias can also result from refusals to participate. Enrolment in CCS was limited to patients who lived within 50 miles of the hospital, the purpose being to include only persons from the population of individuals within that area who would have sought care in the study hospitals and to exclude referrals from outside that area.[62–64] Nonetheless, referral patterns for different cancers or non-malignant illnesses may have differed. Participation rates of targeted patients were high in CCS, but nonetheless many patients were not targeted for interview because they were occupied with tests or visitors or because budgetary constraints limited the number of interviews that could be conducted.

To assess for the potential for selection bias, we may select two control groups, one comprising patients with other cancers judged to be unrelated to the exposure and the second comprising patients with non-malignant conditions thought to be unrelated to the exposure. The selection of the second group guards not just against the possibility that the exposure may cause all cancers, but against the possibility of different referral patterns for patients with malignant and non-malignant illnesses. Uniformity of the exposure of interest across the two control groups suggests the absence of selection bias.

As another check for bias, a disease unrelated to the drug exposure at issue can be included in the assessment of the relation of that drug to the outcome of interest; for example when we assessed acetaminophen in relation to risk of transitional cell cancers, we assessed renal cell cancer as well because this cancer had not been associated with acetaminophen use.[65]

Reporting bias

Medication exposure data based on complete and accurate records would be optimal, with the caveat that people sometimes do not fill prescriptions or take the drugs that have been prescribed. Validation studies of self-reported prescription drug use are difficult in the US because medications may be obtained from multiple sources and records may be absent. Validation of non-prescription drug use or dietary supplement use is infeasible because of the absence of records. A review of validation studies[25]

concluded that reporting of prescription drug use is influenced by the type of medication and drug use patterns (e.g., better reporting for chronically used prescriptions) and by the design of data collection. Because recent or long-term use is likely to be best remembered,[25–30,66–70] we generally place greatest reliance on associations with these categories of use.

The possibility of differential reporting in CCS was reduced by using the same structured interview and same interview setting for cases and controls. Asking about 43 indications for drug use and drug classes served to mask hypotheses about specific drugs. Control patients with conditions serious enough to warrant hospital admission were probably just as likely to carefully search their memories as case patients admitted for cancer.

Exposure misclassification

Studies that focus on a few drugs of interest can obtain more complete information than was collected in CCS, which collected data on all drugs used. Efforts to reduce reporting bias have been described above. Non-differential underreporting of drug use in CCS will have weakened associations, although even underreporting as high as 30% among cases and controls will have resulted in a relatively small shift in the odds ratio estimate toward the null.[71] On the other hand, differential misclassification could have resulted in either increases or decreases in the odds ratios. The ultimate test of validity of CCS results is confirmation in well-conducted studies that use different methods.

Particular applications

As described in this chapter, CCS has the capability to assess the relation of use of prescription and non-prescription medications to risk of a wide range of outcomes. Outcomes assessed to date, and that can be assessed further as new hypotheses or concerns arise, include breast cancer, choriocarcinoma, cholecystitis, endometrial cancer, large bowel cancer and other gastrointestinal cancers, liver cancer, lung cancer, melanoma, non-Hodgkin's lymphoma, ovarian cancer, pelvic inflammatory disease, prostate cancer, and venous thromboembolism (see Appendix). As shown in Table 19.1, there are sufficient numbers of cases of relatively rare outcomes, such as cancers of the testis, pancreas, and esophagus, for informative assessments of drug–disease associations. As also noted, CCS is particularly well suited for assessing associations of diseases with non-prescription medications and medication classes, such as acetaminophen, aspirin, and dietary supplements, for which exposure data can be ascertained only through reporting by the user. Non-drug exposures assessed have included the tar and nicotine content of cigarettes, menthol cigarette smoking, alcohol consumption, coffee consumption, and vasectomy.

The future

Because data collection ceased in CCS in 2009, the database will be useful in the future only for the assessment of medications already in use before data collection ended.

Acknowledgments

CCS was originated in 1975 by Dr Samuel Shapiro and the late Dr Dennis Slone. CCS was supported by the US Food and Drug Administration (cooperative agreements U01 FD 01222-01 and FD-U-000082 and contracts 223-76-3016, 223-80-3001, 226-82-0007) and the National Cancer Institute (grant CA45762). Support for specific data analyses was provided by various pharmaceutical companies, which are acknowledged in the papers that relied on their support.

Appendix: case–control surveillance publications

Rosenberg L, Shapiro S, Kaufman DW, Slone D, Miettinen OS, Stolley PD. Patterns and determinants of conjugated estrogen use. *Am J Epidemiol* 1979; **109**: 676–86.

Kaufman DW, Shapiro S, Rosenberg L, Monson RR, Miettinen OS, Stolley PD, *et al.* Intrauterine contraceptive device use and pelvic inflammatory disease. *Am J Obstet Gynecol* 1980; **136**: 159–62.

Kaufman DW, Slone D, Rosenberg L, Miettinen OS, Shapiro S. Cigarette smoking and age at natural menopause. *Am J Pub Health* 1980; **70**: 420–2.

Shapiro S, Kaufman DW, Slone D, Rosenberg L, Miettinen OS, Stolley PD, *et al.* Recent and past use of conjugated estrogens in relation to adenocarcinoma of the endometrium. *N Engl J Med* 1980; **303**: 485–9.

Rosenberg L, Shapiro S, Slone D, Kaufman DW, Miettinen OS, Stolley PD. Thiazides and acute cholecystitis. *N Engl J Med* 1980; **303**: 546–8.

Shapiro S, Kaufman DW, Slone D, Rosenberg L, Miettinen OS, Stolley PD, *et al.* Use of thyroid supplements in relation to breast cancer. *JAMA* 1980; **244**: 1685–7.

Kaufman DW, Shapiro S, Slone D, Rosenberg L, Miettinen OS, Stolley PD, *et al.* Decreased risk of endometrial cancer among oral contraceptive users. *N Engl J Med* 1980; **303**: 1045–7.

Rosenberg L, Slone D, Shapiro S, Kaufman DW, Helmrich SP, Miettinen OS, *et al.* Breast cancer and alcoholic-beverage consumption. *Lancet* 1982; **i**: 267–70.

Kaufman DW, Shapiro S, Slone D, Rosenberg L, Helmrich SP, Miettinen OS, *et al.* Diazepam and the risk of breast cancer. *Lancet* 1982; **i**: 537–9.

Rosenberg L, Shapiro S, Slone D, Kaufman DW, Helmrich SP, Miettinen OS, *et al.* Epithelial ovarian cancer and the use of combination oral contraceptives. *JAMA* 1982; **247**: 3210–12.

Helmrich SP, Slone D, Shapiro S, Rosenberg L, Kaufman DW, Bain C, *et al.* Risk factors for breast cancer. *Am J Epidemiol* 1983; **117**: 35–45.

Kaufman DW, Watson J, Rosenberg L, Helmrich SP, Miller DR, Miettinen OS, *et al.* The effect of different types of intrauterine device upon the risk of pelvic inflammatory disease. *JAMA* 1983; **250**: 759–62.

Rosenberg L, Schwingl PJ, Kaufman DW, Miller DR, Helmrich SP, Stolley PD, *et al.* Breast cancer and cigarette smoking. *N Engl J Med* 1984; **310**: 92–4.

Rosenberg L, Miller DR, Kaufman DW, Helmrich SP, Stolley PD, Schottenfeld D, *et al.* Breast cancer and oral contraceptive use. *Am J Epidemiol* 1984; **119**: 167–76.

Helmrich SP, Rosenberg L, Kaufman DW, Miller DR, Schottenfeld D, Stolley PD, *et al.* Lack of an elevated risk of malignant melanoma in relation to oral contraceptive use. *J Natl Cancer Inst* 1984; **72**: 617–20.

Shapiro S, Parsells JL, Rosenberg L, Kaufman DW, Stolley PD, Schottenfeld D. Risk of breast cancer in relation to the use of rauwolfia alkaloids. *Eur J Clin Pharmacol* 1984; **26**: 143–6.

Kaufman DW, Miller DR, Rosenberg L, Helmrich SP, Stolley PD, Schottenfeld D, *et al.* Noncontraceptive estrogen use and the risk of breast cancer. *JAMA* 1984; **252**: 63–7.

Rosenberg L, Miller DR, Helmrich SP, Kaufman DW, Shapiro S. Breast cancer and coffee drinking. *Banbury Report 17: Coffee and Health*. Cold Spring Harbor, NY, 1984, pp. 189–95.

Miller DR, Rosenberg L, Helmrich SP, Kaufman DW, Shapiro S. Ovarian cancer and coffee drinking. *Banbury Report 1: Coffee and Health*. Cold Spring Harbor, NY, 1984, pp. 157–65.

Lesko SM, Rosenberg L, Kaufman DW, Helmrich SP, Miller DM, Strom B, *et al.* Cigarette smoking and the risk of endometrial cancer. *N Engl J Med* 1985; **313**: 593–6.

Rosenberg L, Miller DR, Helmrich SP, Kaufman DW, Schottenfeld D, Stolley PD, *et al.* Breast cancer and the consumption of coffee. *Am J Epidemiol* 1985; **122**: 391–9.

Lesko SM, Kaufman DW, Rosenberg L, Helmrich SP, Miller DR, Stolley PD, *et al.* Evidence of increased risk of Crohn's disease in oral contraceptive users. *Gastroenterology* 1985; **89**: 1046–9.

Shapiro S, Parsells JL, Rosenberg L, Kaufman DW, Helmrich SP, Rosenshein N, *et al.* Risk of localized and widespread endometrial cancer in relation to recent and discontinued use of conjugated estrogens. *N Engl J Med* 1985; **313**: 969–72.

Miller DR, Rosenberg L, Kaufman DW, Schottenfeld D, Stolley PD, Shapiro S. Breast cancer risk in relation to early contraceptive use. *Obstet Gynecol* 1986; **68**: 863–8.

Miller DR, Rosenberg L, Kaufman DW, Helmrich SP, Schottenfeld D, Lewis J, *et al.* Epithelial ovarian cancer and coffee drinking. *Int J Epidemiol* 1987; **16**: 13–17.

Helmrich SP, Rosenberg L, Kaufman DW, Strom B, Shapiro S. Venous thromboembolism in relation to oral contraceptive use. *Obstet Gynecol* 1987; **69**: 91–5.

Kaufman DW, Kelly JP, Rosenberg L, Stolley PD, Schottenfeld D, Shapiro S. Hydralazine and breast cancer. *J Natl Cancer Inst* 1987; **78**: 243–6.

Schatzkin A, Palmer JR, Rosenberg L, Helmrich SP, Miller DR, Kaufman DW, *et al.* Risk factors for breast cancer in black women. *J Natl Cancer Inst* 1987; **18**: 213–17.

Rosenberg L, Lesko SM. Cigarette smoking and endometrial cancer. In: Rosenberg MJ, ed. *Smoking and Reproductive Health*. Littleton, MA: PSG, 1987, pp. 160–6.

Levy M, Miller D, Kaufman D, Siskind V, Schwingl P, Rosenberg L, *et al.* Major upper gastrointestinal bleeding and the use of aspirin and other nonnarcotic analgesics. *Arch Intern Med* 1988; **148**: 281–5.

Rosenberg L, Palmer JR, Kaufman DW, Strom BL, Schottenfeld D, Shapiro S. Breast cancer in relation to the occurrence and timing of induced and spontaneous abortion. *Am J Epidemiol* 1988; **127**: 981–9.

Miller DR, Rosenberg L, Kaufman DW, Stolley P, Warshauer ME, Shapiro S. Breast cancer before age 45 and oral contraceptive use: new findings. *Am J Epidemiol* 1989; **129**: 269–80.

Kaufman DW, Palmer JR, Rosenberg L, Stolley P, Warshauer E, Shapiro S. Tar content of cigarettes in relation to lung cancer. *Am J Epidemiol* 1989; **129**: 703–11.

Kaufman DW, Kelly JP, Rosenberg L, Stolley PD, Warshauer ME, Shapiro S. Hydralazine use in relation to cancers of the lung, colon, and rectum. *Eur J Clin Pharmacol* 1989; **36**: 259–64.

Rosenberg L, Palmer JR, Shapiro S. Gestational trophoblastic disease and oral-contraceptive use. *Am J Obstet Gynecol* 1989; **161**: 1087–8.

Rosenberg L, Werler MM, Palmer JR, Kaufman DW, Warshauer ME, Stolley PD, *et al.* The risks of cancers of the colon and rectum in relation to coffee consumption. *Am J Epidemiol* 1989; **130**: 895–903.

Palmer JR, Rosenberg L, Kaufman DW, Warshauer ME, Stolley P, Shapiro S. Oral contraceptive use and liver cancer. *Am J Epidemiol* 1989; **130**: 878–82.

Kaufman DW, Kelly JP, Welch WR, Rosenberg L, Stolley PD, Warshauer ME, *et al.* Noncontraceptive estrogen use and epithelial ovarian cancer. *Am J Epidemiol* 1989; **130**: 1142–51.

Kelly JP, Rosenberg L, Kaufman DW, Shapiro S. Reliability of personal interview data in a hospital-based case–control study. *Am J Epidemiol* 1990; **131**: 79–90.

Kaufman DW, Werler MM, Palmer JR, Rosenberg L, Stolley PD, Warshauer ME, *et al.* Diazepam use in relation to breast cancer: results from two case–control studies. *Am J Epidemiol* 1990; **131**: 483–90.

Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Stolley PD, Shapiro S. Vasectomy and the risk of prostate cancer. *Am J Epidemiol* 1990; **132**: 1051–5.

Austin KL, Palmer JR, Seddon JM, Glynn RJ, Rosenberg L, Gragoudas ES, *et al.* Case control study of idiopathic retinal detachment. *Int J Epidemiol* 1990; **19**: 1045–9.

Miller DR. Breast cancer and early oral contraceptive use: studies from the Slone Epidemiology Unit. In: Mann RD, ed. *Oral Contraceptives and Breast Cancer*. Park Ridge, NJ: Parthenon, 1990, pp. 137–57.

Rosenberg L, Palmer JR, Shapiro S. Oral contraceptive use in relation to the risk of breast cancer: the data of Slone Epidemiology Unit in women aged 45–59 years. In: Mann RD, ed. *Oral Contraceptives and Breast Cancer*. Park Ridge, NJ: Parthenon, 1990, pp. 159–68.

Palmer JR. Oral contraceptive use and gestational choriocarcinoma. *Cancer Detect Prev* 1990; **15**: 45–8.

Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Stolley PD, Shapiro S. A hypothesis: NSAIDs reduce the incidence of large bowel cancer. *J Natl Cancer Inst* 1991; **83**: 355–8.

Lesko SM, Rosenberg L, Kaufman DW, Stolley PD, Warshauer ME, Lewis JL, *et al.* Endometrial cancer and age at last delivery: evidence of an association. *Am J Epidemiol* 1991; **133**: 554–9.

Palmer JR, Rosenberg L, Clarke EA, Stolley PD, Warshauer ME, Zauber AG, *et al.* Breast cancer and cigarette smoking: a hypothesis. *Am J Epidemiol* 1991; **134**: 1–13.

Johannes CB, Kaufman DW, Rosenberg L, Palmer JR, Stolley PD, Lewis JL Jr, *et al.* Side of origin of epithelial ovarian cancer. *BMJ* 1992; **304**: 27–8.

Kaufman DW, Palmer JR, de Mouzon J, Rosenberg L, Stolley PD, Warshauer E, *et al.* Estrogen replacement therapy and the risk of breast cancer: results from the case–control surveillance study. *Am J Epidemiol* 1991; **134**: 1375–85.

Palmer JR, Rosenberg L, Strom BL, Harlap S, Zauber AG, Warshauer ME, *et al.* Oral contraceptive use and risk of cutaneous malignant melanoma. *Cancer Causes Control* 1992; **3**: 547–54.

Kaufman DW, Palmer JR, Rosenberg L, Shapiro S. Oestrogen replacement therapy and breast cancer: new results from the Slone Epidemiology Unit's Case–Control Surveillance Study. In: Mann RD, ed. *Hormone Replacement Therapy and Breast Cancer Risk*. Park Ridge, NJ: Parthenon, 1992, pp. 173–84.

Shapiro S. Case–control surveillance. In: Strom BL, ed. *Pharmacoepidemiology*, 2nd edn. Chichester: John Wiley & Sons, 1994, pp. 301–22.

Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Lewis JL Jr, Strom BL, *et al.* A case–control study of oral contraceptive use and invasive epithelial ovarian cancer. *Am J Epidemiol* 1994; **139**: 654–61.

Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Strom BL, Harlap S, *et al.* The relation of vasectomy to the risk of cancer. *Am J Epidemiol* 1994; **140**: 431–8.

Rosenberg L. NSAIDs and cancer. *Prev Med* 1995; **24**: 107–9.

Palmer JR, Rosenberg L, Harlap S, Strom BL, Warshauer ME, Zauber AG, *et al.* Adult height and risk of breast

cancer among US black women. *Am J Epidemiol* 1995; **141**: 845–9.

Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Strom BL, Harlap S, *et al.* The relation of benzodiazepine use to the risk of selected cancers: breast, large bowel, malignant melanoma, lung, endometrium, ovary, non-Hodgkin's lymphoma, testis, Hodgkin's disease, thyroid and liver. *Am J Epidemiol* 1995; **141**: 1153–60.

Palmer JR, Rosenberg L, Rao RS, Strom BL, Warshauer ME, Harlap S, Zauber A, *et al.* Oral contraceptive use and breast cancer risk among African-American women. *Cancer Causes Control* 1995; **6**: 321–3.

Rosenberg L, Palmer JR, Rao RS, Zauber AG, Strom BL, Warshauer ME, *et al.* Case–control study of oral contraceptive use and risk of breast cancer. *Am J Epidemiol* 1996; **143**: 25–37.

Zhang Y, Rosenberg L, Colton T, Couples A, Palmer JR, Strom BL, *et al.* Adult height and risk of breast cancer among white women in a case–control study. *Am J Epidemiol* 1996; **143**: 1123–8.

Palmer JR, Rosenberg L, Rao RS, Zauber A, Strom BL, Warshauer ME, *et al.* Induced and spontaneous abortion in relation to risk of breast cancer (United States). *Cancer Causes Control* 1997; **8**: 841–9.

Rosenberg L, Rao RS, Palmer JP, Strom BL, Zauber AG, Warshauer ME, *et al.* Transitional cell cancer of the urinary tract and renal cell cancer in relation to acetaminophen use. *Cancer Causes Control* 1998; **9**: 83–8.

Rosenberg L, Rao RS, Palmer JR, Strom BL, Stolley PD, Zauber AG, *et al.* Calcium channel blockers and the risk of cancer. *JAMA* 1998; **279**: 1000–4.

Rosenberg L, Stephenson WP, Rao RS, Palmer JR, Strom BL, Shapiro S. The diagnosis of renal cell cancer in relation to hypertension (United States). *Cancer Causes Control* 1998; **9**: 611–14.

Shapiro S. Case–control surveillance. In: Strom BL, ed. *Pharmacoepidemiology*, 3rd edn. Chichester: John Wiley & Sons, 2000, pp. 209–30.

Coogan PF, Rao RS, Rosenberg L, Palmer JR, Strom BL, Zauber AG, *et al.* The relationship of nonsteroidal anti-inflammatory drug use to the risk of breast cancer. *Prev Med* 1999; **29**: 72–6.

Kelly JP, Rosenberg L, Palmer JR, Rao RS, Strom BL, Stolley PD, *et al.* Risk of breast cancer according to use of antidepressants, phenothiazines, and antihistamines. *Am J Epidemiol* 1999; **150**: 675–82.

Coogan PF, Rosenberg L, Palmer JR, Strom BL, Zauber AG, Stolley PD, *et al.* Nonsteroidal anti-inflammatory drugs and risk of digestive cancers at sites other than the large bowel. *Cancer Epidemiol Biomark Prev* 2000; **9**: 119–23.

Coogan PF, Rosenberg L, Zauber AG, Stolley PD, Strom BL, Shapiro S. NSAIDs and risk of colorectal cancer according to presence or absence of family history of disease. *Cancer Causes Control* 2000; **11**: 249–55.

Rosenberg L, Palmer JR, Rao RS, Coogan PF, Strom BL, Zauber AG, *et al.* A case–control study of analgesic use and ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2000; **9**: 933–7.

Coogan PF, Rosenberg L, Palmer JR, Strom BL, Stolley PD, Zauber AG, *et al.* Risk of ovarian cancer according to use of antidepressants, phenothiazines, and benzodiazepines. *Cancer Causes Control* 2000; **11**: 839–45.

Coogan PF, Rosenberg L, Palmer JR, Strom BL, Zauber AG, Stolley PD, *et al.* Phenolphthalein laxatives and risk of cancer. *J Natl Cancer Inst* 2000; **92**: 1943–4.

Kaufman DW, Rosenberg L, Mitchell AA. Signal generation and clarification: use of case–control data. *Pharmacoepidemiol Drug Saf* 2001; **10**: 197–203.

Coogan PF, Rosenberg L, Palmer JR, Strom BL, Zauber AG, Shapiro S. Statin use and the risk of breast and prostate cancer. *Epidemiology* 2002; **13**: 262–7.

Blackman JA, Coogan PF, Rosenberg L, Strom BL, Zauber AG, Palmer JR, *et al.* Estrogen replacement therapy and risk of lung cancer by cell type. *Pharmacoepidemiol Drug Saf* 2002; **11**: 561–7.

Coogan P, Rosenberg L. Response to Etminan letter Re: Statin use and the risk of breast and prostate cancer [letter]. *Epidemiology* 2002; **13**: 607–8.

Brooks DR, Palmer JR, Strom BL, Rosenberg L. Menthol cigarettes and risk of lung cancer. *Am J Epidemiol* 2003; **158**: 609–16.

Zhang Y, Coogan PF, Palmer JR, Strom BL, Rosenberg L. Cigarette smoking and increased risk of mucinous epithelial ovarian cancer. *Am J Epidemiol* 2004; **159**: 133–9.

Rosenberg L, Coogan PF, Palmer JR. Case-control surveillance. In: Strom BL, ed. *Pharmacoepidemiology*, 4th edn. New York: Wiley, 2005, 185–202.

Coogan PF, Zhang Y, Palmer JR, Strom BL, Rosenberg L. Cimetidine and other H_2 blocker use in relation to risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2005; **14**: 1012–15.

Zhang Y, Coogan PF, Palmer JR, Strom BL, Rosenberg L. Use of non-steroidal anti-inflammatory drugs and risk of breast cancer: The case-control surveillance study revisited. *Am J Epidemiol* 2005; **162**: 165–70.

Coogan PF, Palmer JR, Strom BL, Rosenberg L. Use of selective serotonin reuptake inhibitors and the risk of breast cancer. *Am J Epidemiol* 2005; **162**: 835–8.

Zhang Y, Coogan PF, Palmer JR, Strom BL, Rosenberg L. Risk of non-Hodgkin's lymphoma and use of non-steroidal anti-inflammatory drugs. *Cancer Detect Prev* 2006; **30**: 99–101.

Rosenberg L, Coogan PF, Palmer JR. Case-control surveillance. In: Strom BL, Kimmel SE, eds. *Textbook of Pharmacoepidemiology* , 5th edn. New York: Wiley, 2006, 137–49.

Coogan PF, Rosenberg L, Strom BL. Statin use and the risk of 10 cancers. *Epidemiology* 2007; **18**: 213–19.

Coogan PF, Strom BL, Rosenberg L. SSRI use and breast cancer risk by hormone receptor status. *Breast Cancer Res Treat* 2007; **109**: 527–31.

Collaborative Group on Epidemiological Studies of Ovarian Cancer (Coogan PC, Rosenberg L). Ovarian cancer and oral contraceptives: Collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. *Lancet* 2008; **371**: 303–14.

Coogan PF, Strom BL, Rosenberg L. Diuretic use and the risk of breast cancer. *J Hum Hypertens* 2009; **23**: 216–19.

Rosenberg L, Zhang Y, Coogan PF, Palmer JR. A Case-control study of oral contraceptive use and incident breast cancer. *Am J Epidemiol* 2009; **169**: 473–9.

Zhang Y, Coogan PF, Palmer JR, Strom BL, Rosenberg L. Vitamin and mineral use and risk of prostate cancer: The Case-Control Surveillance Study. *Cancer Causes Control* 2009; **20**: 691–8.

Coogan PF, Strom BL, Rosenberg L. Antidepressant use and colorectal cancer risk. *Pharmacoepidemiol Drug Saf* 2009; **18**: 1111–14.

Zhang Y, Coogan PF, Palmer JR, Strom BL, Rosenberg L. A case-control study of reproductive factors, female hormone use, and risk of pancreatic cancer. *Cancer Causes Control* 2009; **21**: 473–8.

Rosenberg L, Zhang Y, Coogan PF, Strom BL, Palmer JR. Response to Shapiro letter Re: A case-control study of oral contraceptive use and incident breast cancer [letter]. *Am J Epidemiol* 2009; **170**: 803–4.

Coogan PF, Kelly JP, Strom B, Rosenberg L. Statin and NSAID use and prostate cancer risk. *Pharmacoepidemiol Drug Saf* 2010; **19**: 752–5.

References

1. Brewer T, Colditz G. Postmarketing surveillance and adverse drug reactions: current perspectives and future needs. *JAMA* 1999; **281**: 824–9.
2. Strom BL. Overview of automated databases in pharmacoepidemiology. In: Strom BL, ed. *Pharmacoepidemiology*, 4th edn. New York: Wiley, 2005, pp. 219–22.
3. Soller RW. OTCS 2000: Achievements and challenges. *Drug Inf J* 2000; **34**: 693–701.
4. Francesco International. *Potential Government High Priority Therapeutic Areas for OTC*. SWITCH® Newsletter, SwitchTrends. http://www.rxtootcswitch.com (accessed Jan. 2004).
5. Kaufman DW, Kelley JP, Rosenberg L, Anderson TE, Mitchell AA. Recent patterns of medication use in the US: the Slone Survey. *JAMA* 2002; **287**: 337–44.
6. Kessler DA. Cancer and herbs. *N Engl J Med* 2000; **342**: 1762–3.
7. Newall CA, Anderson LA, Philpson JD. *Herbal Medicine: A Guide for Healthcare Professionals*. London: The Pharmaceutical Press, 1966.
8. Nortier JL, Martinez MC, Schmeiser HH, Arlt VM, Bieler A, Petein M, *et al.* Urothelial carcinoma associated with the use of a Chinese herb (*Aristolochia fangchi*). *N Engl J Med* 2000; **342**: 1686–92.
9. Czeczot H, Tudek B, Kusztelak J, Szynczyk T, Dobrowolska B, Glinkowska G, *et al.* Isolation and studies of mutagenic activity in the Ames test of flavonoids naturally occurring in medical herbs. *Mutat Res* 1990; **240**: 209–16.
10. Kapadia GJ, Chung EB, Ghosh B, Shukla YN, Besak SP, Morton JF, *et al.* Carcinogenicity of some folk medicinal herbs in rats. *J Natl Cancer Inst* 1978; **60**: 683–6.
11. Shin HR, Kim JY, Yun TK, Morgan G, Vainio H. The cancer-preventive potential of Panax ginseng: a review of human and experimental evidence. *Cancer Causes Control* 2000; **11**: 565–76.
12. Eagon PK, Elm MS, Hunter DS. Medicinal herbs: modulation of estrogen action. *Era of Hope Meeting for the Department of Defense Breast Cancer Research Program*; June 8–11, 2000, Atlanta, Georgia.
13. Dixon-Shanies D, Shaich N. Growth inhibition of human breast cancer cells by herbs and phytoestrogens. *Oncol Rep* 1999; **6**: 1383–7.
14. Gurley BJ, Gardner SF, Hubbard MA. Clinical assessment of potential cytochrome P450-mediated herb–drug interactions. *AAPS Annual Meeting and Exposition*, presentation #3460; Oct 9–Nov 2, 2000, Indianapolis, Indiana.
15. Budzinski JW, Foster BC, Vandenhoek S, Arnason JT. An in vitro evaluation of human cytochrome P450 3A4 inhibition by selected commercial herbal extracts and tinctures. *Phytomedicine* 2000; **7**: 273–82.

16. Rosenberg L, Shapiro S, Slone D, Kaufman DW, Miettinen OS, Stolley PD. Thiazides and acute cholecystitis. *N Engl J Med* 1980; **303**: 546–8.
17. Kaufman DW, Shapiro S, Rosenberg L, Monson RR, Miettinen OS, Stolley PD, *et al.* Intrauterine contraceptive device use and pelvic inflammatory disease. *Am J Obstet Gynecol* 1980; **136**: 159–62.
18. Rothman KJ. *Modern Epidemiology*. Boston, MA: Little Brown, 1986, p. 60.
19. Schlesselman SS. *Case–Control Studies: Design, Conduct, Analysis*. New York: Oxford University Press, 1982.
20. Miettinen OS. Valid sampling in the context of a base secondary to the cases. In: *Theoretical Epidemiology. Principles of Occurrence Research in Medicine*. New York: John Wiley & Sons, 1985, pp. 79–82.
21. Pike MC, Peters RK, Cozen W, Probst-Hensch NM, Felix JC, Wan PC, *et al.* Estrogen-progestin replacement therapy and endometrial cancer. *J Natl Cancer Instit* 1997; **89**: 1110–6
22. Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Stolley PD, Shapiro S. A hypothesis: NSAIDs reduce the incidence of large bowel cancer. *J Natl Cancer Inst* 1991; **83**: 355–8.
23. Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Strom BL, Harlap S, *et al.* The relation of vasectomy to the risk of cancer. *Am J Epidemiol* 1994; **140**: 431–8.
24. Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Lewis JL Jr, Strom BL, *et al.* A case–control study of oral contraceptive use and invasive epithelial ovarian cancer. *Am J Epidemiol* 1994; **139**: 654–61.
25. West SL, Strom BL. Validity of pharmacoepidemiology drug and diagnosis data. In: Strom BL, ed., *Pharmacoepidemiology*, 2nd edn. Chichester: John Wiley & Sons, 1994; pp. 199–216.
26. Nischan P, Ebeling K, Thomas DB, Hirsch U. Comparison of recalled and validated oral contraceptive histories. *Am J Epidemiol* 1993; **138**: 697–703.
27. Glass R, Johnson B, Vessey M. Accuracy of recall of histories of oral contraceptive use. *Br J Prev Soc Med* 1974; **28**: 273–5.
28. 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.
29. Stolley PD, Tonascia JA, Sartwell PE, Tockman MS, Tonascia S, Rutledge A, *et al.* Agreement rates between oral contraceptive users and prescribers in relation to drug use histories. *Am J Epidemiol* 1978; **107**: 226–35.
30. van Leeuwen FE, van Duijn CM, Camps MH, Kempers BA, Mentjens MF, Mulder HB, *et al.* Agreement between oral contraceptive users and prescribers: implications for case–control studies. *Contraception* 1992; **45**: 399–408.
31. Kaufman DW, Palmer JR, de Mouzon J, Rosenberg L, Stolley PD, Warshauer E, *et al.* Estrogen replacement therapy and the risk of breast cancer: results from the case–control surveillance study. *Am J Epidemiol* 1991; **134**: 1375–85.
32. Thun MJ, Namboodiri MM, Heath CW Jr. Aspirin use and reduced risk of fatal colon cancer. *N Engl J Med* 1991; **325**: 1593–6.
33. Gann PH, Manson JE, Glynn RJ, Buring JE, Hennekens CH. Low-dose aspirin and incidence of colorectal tumors in a randomized trial. *J Natl Cancer Inst* 1993; **35**: 1220–4.
34. Giovannucci E, Egan KM, Hunter DJ, Stampfer MJ, Colditz GA, Willett WC, *et al.* Aspirin and the risk of colorectal cancer in women. *N Engl J Med* 1995; **333**: 609–14.
35. Thun MS, Namboodiri MM, Calle EE, Flanders MD, Heath CW Jr. Aspirin use and risk of fatal cancer. *Cancer Res* 1993; **53**: 1322–7.
36. Rosenberg L, Louik C, Shapiro S. Nonsteroidal antiinflammatory drug use and reduced risk of large bowel carcinoma. *Cancer* 1998; **82**: 2326–33.
37. Rosenberg L, Palmer JR, Shapiro S. Gestational trophoblastic disease and oral-contraceptive use. *Am J Obstet Gynecol* 1989; **161**: 1087–8.
38. Lesko SM, Kaufman DW, Rosenberg L, Helmrich SP, Miller DR, Stolley PD, *et al.* Evidence of increased risk of Crohn's disease in oral contraceptive users. *Gastroenterology* 1985; **89**: 1046–9.
39. Palmer JR, Driscoll SD, Rosenberg L, Berkowitz RS, Lurain JR, Soper J, *et al.* Oral contraceptive use and risk of gestational trophoblastic tumors. *J Natl Cancer Inst* 1999; **91**: 635–40.
40. Boyko EJ, Theis MK, Vaughan TL, Nicol-Blades B. Increased risk of inflammatory bowel disease associated with oral contraceptive use. *Am J Epidemiol* 1994; **140**: 268–78.
41. Rosenberg L, Slone D, Shapiro S, Kaufman DR, Helmrich SP, Miettinen OS, *et al.* Breast cancer and alcoholic-beverage consumption. *Lancet* 1982; **i**: 267–70.
42. Rosenberg L, Metzger LS, Palmer JR. Epidemiology of breast cancer. Alcohol consumption and risk of breast cancer: a review of the evidence. *Epidemiol Rev* 1993; **15**: 133–44.

43. Willett W, Stampfer MJ, Colditz GA, Rosner BA, Hennekens CH, Speizer FE. Moderate alcohol consumption and the risk of breast cancer. *N Engl J Med* 1987; **316**: 1174–80.
44. Schatzkin A, Jones DY, Hoover RN, Taylor PR, Brinton LA, Ziegler RG, *et al.* Alcohol consumption and breast cancer in the epidemiologic follow-up study of the First National Health and Nutrition Examination Survey. *N Engl J Med* 1987; **316**: 1169–73.
45. Palmer JR, Rosenberg L, Kaufman DW, Warshauer ME, Stolley P, Shapiro S. Oral contraceptive use and liver cancer. *Am J Epidemiol* 1989; **130**: 878–82.
46. Rosenberg L, Miller DR, Kaufman DW, Helmrich SP, Stolley PD, Schottenfeld D, *et al.* Breast cancer and oral-contraceptive use. *Am J Epidemiol* 1984; **119**: 167–76.
47. Rosenberg L, Palmer JR, Rao RS, Zauber AG, Strom BL, Warshauer ME, *et al.* Case–control study of oral contraceptive use and risk of breast cancer. *Am J Epidemiol* 1996; **143**: 25–37.
48. Rosenberg L, Zhang Y, Coogan PF, Palmer JR. A case-control study of oral contraceptive use and incident breast cancer. *Am J Epidemiol* 2009; **169**: 473–9.
49. Shapiro S, Kelly JP, Rosenberg L, Kaufman DW, Helmrich SP, Rosenshein NB, *et al.* Risk of localized and widespread endometrial cancer in relation to recent and discontinued use of conjugated estrogens. *N Engl J Med* 1985; **313**: 969–72.
50. Kaufman DW, Shapiro S, Slone D, Rosenberg L, Miettinen OS, Stolley PD, *et al.* Decreased risk of endometrial cancer among oral contraceptive users. *N Engl J Med* 1980; **303**: 1045–7.
51. Coogan PF, Rosenberg L, Palmer JR, Strom BL, Zauber AG, Stolley PD, *et al.* Nonsteroidal anti-inflammatory drugs and risk of digestive cancers at sites other than the large bowel. *Cancer Epidemiol Biomarkers Prev* 2000; **9**: 119–23.
52. National Toxicology Program. Technical report on the toxicology and carcinogenesis studies of phenolphthalein (CAS no. 77-09-8). In: *F344/N Rats and B6C3F, Mice (Feed Studies) NTP TR 465*. NIH Publication no. 95-3390, 1995.
53. Coogan PF, Rosenberg L, Palmer JR, Strom BL, Zauber AG, Stolley PD, *et al.* Phenolphthalein laxatives and risk of cancer. *J Natl Cancer Inst* 2000; **92**: 1943–4.
54. Pahor M, Guralnik JM, Salive ME, Corti M-C, Carbonin P, Havlik RJ. Do calcium channel blockers increase the risk of cancer? *Am J Hypertens* 1996; **9**: 695–9.
55. Pahor M, Guralnik JM, Corti M-C, Salive ME, Cerhan JR, Wallace RB, *et al.* Calcium-channel blockade and incidence of cancer in aged populations. *Lancet* 1996; **348**: 493–7.
56. Rosenberg L, Rao RS, Palmer JR, Strom BL, Stolley PD, Zauber AG, *et al.* Calcium channel blockers and the risk of cancer. *JAMA* 1998; **279**: 1000–4.
57. Kaufman DW, Shapiro S, Slone D, Rosenberg L, Helmrich SP, Miettinen OS, *et al.* Diazepam and the risk of breast cancer. *Lancet* 1982; **i**: 537–9.
58. Kaufman DW, Werler MM, Palmer JR, Rosenberg L, Stolley PD, Warshauer ME, *et al.* Diazepam use in relation to breast cancer: results from two case–control studies. *Am J Epidemiol* 1990; **131**: 483–90.
59. Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Strom BL, Harlap S, *et al.* The relation of benzodiazepine use to the risk of selected cancers: breast, large bowel, malignant melanoma, lung, endometrium, ovary, non-Hodgkin's lymphoma, testis, Hodgkin's disease, thyroid and liver. *Am J Epidemiol* 1995; **141**: 1153–60.
60. Kaufman DW, Kelly JP, Rosenberg L, Stolley PD, Schottenfeld D, Shapiro S. Hydralazine and breast cancer. *J Natl Cancer Inst* 1987; **78**: 243–6.
61. Kaufman DW, Kelly JP, Rosenberg L, Stolley PD, Warshauer ME, Shapiro S. Hydralazine use in relation to cancers of the lung, colon, and rectum. *Eur J Clin Pharmacol* 1989; **36**: 259–64.
62. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case–control studies. I. Principles. *Am J Epidemiol* 1992; **135**: 1019–28.
63. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case–control studies. II. Types of controls. *Am J Epidemiol* 1992; **135**: 1029–41.
64. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case–control studies. III. Design options. *Am J Epidemiol* 1992; **135**: 1042–50.
65. Rosenberg L, Rao RS, Palmer JP, Strom BL, Zauber AG, Warshauer ME, *et al.* Transitional cell cancer of the urinary tract and renal cell cancer in relation to acetaminophen use. *Cancer Causes Control* 1998; **9**: 83–8.
66. Spengler RF, Clarke EA, Woolever CA, Newman AM, Osborn RW. Exogenous estrogens and endometrial cancer: a case–control study and assessment of potential biases. *Am J Epidemiol* 1981; **114**: 497–506.
67. Goodman MT, Nomura AMY, Wilkens LR, Kolonel LN. Agreement between interview information and physician records on history of menopausal estrogen use. *Am J Epidemiol* 1990; **131**: 815–25.

68. Persson I, Bergkvist L, Adami H-O. Reliability of women's histories of climacteric oestrogen treatment assessed by prescription forms. *Int J Epidemiol* 1987; **16**: 222–8.
69. Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 1982; **116**: 114–22.
70. Harlow SD, Linet MS. Agreement between questionnaire data and medical recall: the evidence of accuracy of recall. *Am J Epidemiol* 1989; **129**: 233–48.
71. Rosenberg L, Coogan PF, Palmer JR. Case-control surveillance. In: Strom BL, Kimmel SE, eds. *Textbook of Pharmacoepidemiology*, 5th edn. New York: Wiley, 2006, 137–49.

CHAPTER 20
Prescription–Event Monitoring

Deborah Layton and Saad A.W. Shakir
Drug Safety Research Unit, Southampton, UK

Introduction

The thalidomide disaster, which caused phocomelia in nearly 10 000 children whose mothers took thalidomide during pregnancy, was the stimulus for the establishment of systems to monitor suspected adverse drug reactions (ADRs) and the development of modern pharmacovigilance.[1] Pharmacovigilance is concerned with detection, assessment, and prevention of adverse effects or any other possible drug-related problems; its ultimate goal is to achieve rational and safe therapeutic decisions in clinical practice. It relies on sources of evidence derived from all levels, particularly pharmacoepidemiologic research. The reasons for monitoring postmarketing drug safety were summarized as follows in 1970 in a report of the Committee on Safety of Drugs in the UK (which later became the Committee on Safety of Medicines (CSM); now the Commission on Human Medicines (CHM)):[2]

> 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.

Clinical experience showed that unexpected hazards could occur with old drugs as well as newly licensed drugs and these realizations thus defined the purpose of "postmarketing surveillance". The limitations of premarketing studies in defining the necessary safety profiles of drugs include: sample size, duration, and patient selection (see Chapter 1). Therefore, there has been general agreement for more than 40 years of the importance of postmarketing adverse event monitoring and postmarketing safety studies in providing complementary information on the clinically necessary understanding of the safety of a drug. This has resulted in not only the establishment of voluntary systems for reporting suspected adverse drug reactions (ADRs) (see Chapter 10), but the development of a range of other methods to monitor and study postmarketing drug safety.

Soon after the establishment of spontaneous reporting systems (see Chapter 10), it was recognized that, while such systems have many real advantages for identifying ADRs, particularly rare ADRs, they also have limitations.[3] The theoretical basis for establishing a system to monitor events regardless of relatedness to drug exposure was proposed by Finney in 1965.[4] This, and the limited contribution of the spontaneous reporting system for detecting hazards such as the oculomucocutaneous syndrome with the beta-blocker practolol,[5] led William Inman to set up in 1980 the Postmarketing Drug Surveillance Research Unit, with financial assistance from the Office of the Chief Scientists of the Department of Health Social Security.[6,7] Subsequently the CSM, wishing to consider monitoring the postapproval safety of

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

medicines, established a working party on Adverse Reactions under the chairmanship of David Grahame-Smith. In June 1983 and again in July 1985, this working party reported an appreciation of the need for a prescription-based monitoring process (prescription–event monitoring; PEM) to provide a means of monitoring new drugs destined for widespread, long-term use. The underlying purpose was to extend the safety database of a new drug to at least 10 000 exposed individuals.

Initially part of the Department of Medicine of the University of Southampton, the Unit was reconstituted as a charitable trust in 1986 and its name was changed to the Drug Safety Research Unit (DSRU). The DSRU is an independent, registered, medical, non-profit organization and now operates in association with the University of Portsmouth. PEM is the only national system in England used to monitor the safety of recently marketed medicines available to all primary care physicians (general practitioners; GPs), besides the Yellow Card system. The Yellow Card system is the UK's spontaneous ADR reporting scheme run by the MHRA and the Commission on Human Medicines (CHM) (see Chapter 10). It receives reports of suspected ADRs from health-care professionals and patients for medicines and vaccines. There are important differences in the type of data collected in PEM compared with the Yellow Card system, the most important being that the majority of events reported in PEM 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. PEM provides estimates of common to rare events while the Yellow Card reporting 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 PEM provide complementary information on hazards associated with medicines.

Description

Design and source data

PEM uses an observational cohort design for active surveillance of targeted medicinal products in England. Products that are selected for study by PEM are new medicinal products which are expected to be widely used in general practice, or established products when there is a reason to do so, for example, a new indication or extending usage to a new population. PEM utilizes the structure of the UK National Health Service (NHS), whereby all individuals are registered with general practitioners (GP), who provide primary medical care and act 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.

The identification of individuals who have received a PEM study drug of interest relies on these NHS prescription data. Patients have to get these prescriptions dispensed at pharmacies with an NHS contract. Pharmacists then submit information on prescriptions dispensed for products available for reimbursement through the NHS to a central prescription processing center within the NHS Business Services Authority, known as the NHS Prescription Services (NHSRxS); formerly the Prescription Pricing Division (PPD). Under longstanding arrangements and through secure transmission, the DSRU is provided with electronic copies of all those prescriptions issued throughout England for the drugs being monitored (see also subsection Ethics and confidentiality). Since the NHSRxS only handles the remuneration and reimbursement to dispensing contractors across England, data are not available for Scotland, Wales, and Northern Ireland. The NHSRxS receives remuneration from the DSRU for this service. These data are reconciled with GP identifier records available from the NHS Organisation Data Services (ODS), to obtain prescriber contact details and, with existing records on the DSRU customized PEM database, to ascertain whether the data pertain to an existing

eligible patient already within the DSRU PEM database. It should be noted that all relevant prescriptions are collected, irrespective of whether they are a new or repeat course. These arrangements operate for the length of time necessary for the DSRU 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.

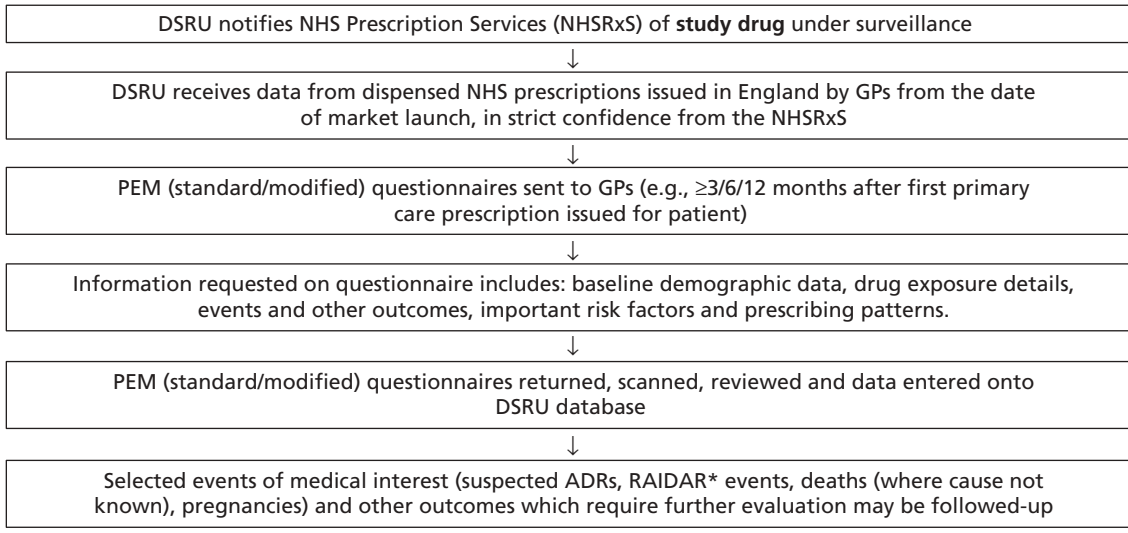
Ethics and confidentiality

PEM studies are conducted according to national and international guidelines for ethical conduct of research involving human subjects.[8–11] Following the principles of good practice, for example as described in the Guidelines for Good Pharmacoepidemiology Practices of the International Society for Pharmacoepidemiology,[12] a full protocol is written for each study to monitor and research the safety of medicines. In addition, 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 (October 2009).[13] 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 NHSRxS 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

The process

Relevant study data are currently collected via a manual process (as summarized in Figure 20.1).



*RAIDAR, rare and iatrogenic adverse reactions.

Figure 20.1 The PEM process.

After an interval of usually 6 months (range 3–12 months) from the date of the first prescription for each eligible patient, PEM questionnaires (standard and customized formats, described below) 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. Historically, PEM questionnaires were commonly referred to as "Green Forms," because of their color. These are intended to be simple in order to expedite speed of data collection in order to enhance surveillance and to encourage response in the interest of drug safety, given there is no remuneration to respondents. The number of 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 in Figure 20.2, the standard "Green Form" PEM questionnaires 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 20.1) 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 pharmacoepidemiologic developments in general, PEM questionnaires have evolved over time to extend the range of data that can be collected on drug utilization patterns as well as important risk factors for selected outcomes of interest which can provide complimentary information on safety issues. Although still limited, the original Green Form simple design was enhanced to include a small number of "additional" questions to facilitate the examination for the entire study cohort of aspects such as confounding by indication, concurrent illnesses, and concomitant medications. For example, the green forms in the PEM studies of the selective cyclo-oxygenase (COX)-2 inhibitors, for example celecoxib, included questions regarding previous history of dyspeptic conditions (Figure 20.2) and the green forms for the PEM studies of PDE5 inhibitors for erectile dysfunction, such as sildenafil, included questions about history of cardiovascular disease.

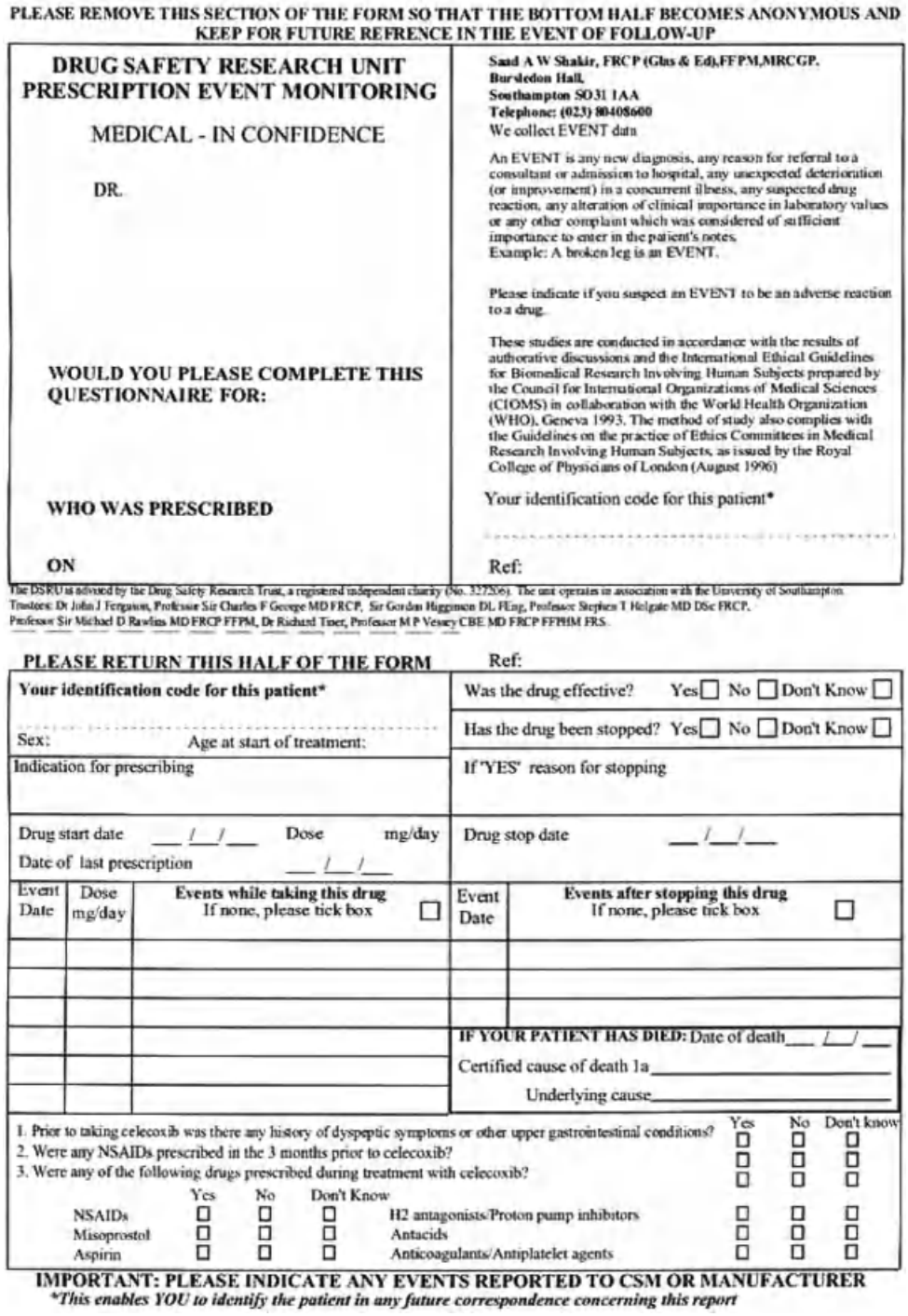
More recently, further expansion and modification of the nature and type of information requested on PEM questionnaires for the whole cohort has led to the adoption of more complex forms; such studies where these forms are used are called "modified PEM" (M-PEM) studies. Customized questionnaires have 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 20.3 illustrates the M-PEM study questionnaire for lumiracoxib. These customized questionnaires are designed to accommodate the increasing need for supplementary information which is relevant to a number of outcomes. This provides the opportunity to investigate and explore a range of additional research questions, such as the identification of subsets of populations potentially at higher risk, the description of factors important in drug utilization and targeted surveillance of specific safety concerns. Because of the increased complexity of these M-PEM questionnaires, GPs receive a small remuneration for returning completed customized forms. Completion of these forms by GPs is voluntary.

Supplemental information

During the course of any study, selected medical events (as described in Table 20.1) 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 GPs using targeted questionnaire(s),

Box 20.1 Definition of an event in prescription–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.



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 20.2 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 20.3 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 (1)	(2)	(3)**	After stopping	Normal range (Units)
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 20.3 (*Continued*)

Table 20.1 Categories of events and outcomes that undergo further evaluation

Medically important* 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 PEM study
Rare and iatrogenic adverse reactions (RAIDAR) events (see Table 20.2)
Any other adverse events deemed to be of medical importance by the DSRU during the PEM study
Specific outcomes associated with the study aims and objectives, for example aspects of prescribing, pre-existing 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

*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."

Table 20.2 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 re-challenge

This list is based on a similar list used by the Medicines and Healthcare products Regulatory Agency (MHRA), UK.

where such information is not provided on the PEM questionnaires. 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. Table 20.2 lists the medically serious events[14] that have been associated with the use of medicines, as compiled by the DSRU. Cases of these events routinely undergo further evaluation (see section Qualitative Evaluation of Important and Medically Important Adverse Events). All pregnancies reported during treatment or within 3 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, provided the reporting GP has supplied a practice identification code for the patient.

Data processing

Each PEM/M-PEM questionnaire returned for each patient is scanned onto the database and electronic copies reviewed by the medical or scientific officer monitoring the study in the DSRU. This initial review aims to identify possible serious ADRs or

events requiring action, for example external communications or expedited follow up. PEM records 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 non-serious.

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 the DSRU Event Dictionary, which has been developed since the inception of PEM in the 1980s. This hierarchical dictionary, which is arranged in a system–organ classification, groups associated doctor summary terms (terminology used by the prescribing physician) under lower-level event terms; similarly, related lower-level event terms are grouped under a broader term (higher-level term). As of July 2010, the dictionary contains 17 593 doctor summary terms (as near as possible to the term used by the reporting doctor, e.g., crescendo angina) and 1939 lower-level terms mapped to 1308 higher-level terms, within 32 system–organ classes. Selected attributes are linked to selected data, for example an event is coded as an ADR if the GP specified that the event is attributable to a drug (either the study drug, or another drug taken during the study observation period); if the ADR has been reported to the CHM or marketing authorization holder; if the event had a fatal outcome; or if the event was a reason for stopping. The DSRU is currently implementing the integration of the Medical Dictionary for Regulatory Activities (MedDRA) dictionary to enhance the description of adverse events in parallel to the use of its own dictionary.

Good clinical data management is a high priority. The DSRU has a set of rules and processes associated with the conduct of PEM 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 diagnose their cause, an effort supported by customized software with objective, standardized logic checks.

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 PEM, the sample size of 10 000 exposed patients has been 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 three" (see Chapter 4), it follows that the larger the sample size, the rarer the event that can be detected. Through aiming to collect a sample size of 10 000 patients in PEM, this allows one to be 95% certain that any events not observed occur less often than 1 in 3333 cases (incidence <0.0003).[15]

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).[16] If the background rate is known and there is an *a priori* 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%.[17] (See Chapter 4 for more details.)

Because of the customized nature of M-PEM 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 best reflecting 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 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 studies that have been undertaken to date, the sample size has been smaller than the 10 000 required for "standard" PEM studies.

The duration of any study is dependent on the level of prescribing of the study drug by GPs in England. Interim analyses are usually undertaken at prespecified milestones (e.g., annually), or defined sample sizes (e.g., 2500 patients in a standard PEM study) 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 109 PEM studies to date with a median cohort size of 11 680 patients (interquartile range: 8670 to 13 632). A wide range of drugs have been studied 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 non-steroidal anti-inflammatory drugs (including selective COX-2 inhibitors), several antibiotics, and antiviral agents have been studied.

Data analysis

Signal detection and evaluation is the primary concern of pharmacovigilance. Several methods are applied for signal detection in PEM, both qualitative and quantitative, not only to look for new unexpected adverse reactions but also for further information regarding expected drug–adverse events associations of interest that might affect the benefit : risk balance of a drug.

Box 20.2 Points for consideration in evaluation of reported events

- The temporal relationship (time to onset)
- The clinical and pathological 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 postmarketing 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 de-challenge or dose reduction
- The effect of re-challenge 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

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 20.2 and Chapter 33). An example of a safety signal generated in PEM as a result of careful clinical evaluation is the visual field defects in patients receiving long-term treatment with the antiepileptic drug vigabatrin.[18]

Because of the epidemiologic nature of the design of PEM, any inferences on drug-relatedness will be made on aggregate basis at study milestones, that is when the interim and final reports are written. Such aggregate analyses can help formulate possible hypotheses, which then require further analytic study. PEM 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.[19]

Quantitative analysis of events

The PEM method provides a numerator (the number of reports of an event) and also denomina-

tors in terms of the number of patients and the number of patient-months of exposure to the drug; all 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 "higher-level" event terms from the DSRU dictionary. As described in Chapter 10, the trend of reports after starting treatment may be informative: pharmacologic 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 when compared with that expected from the Summary of Product Characteristics (SPC). An example is gynecomastia with finasteride (Case study 20.1).[20]

Analysis by event rates (incidence densities)

PEM takes advantage of the information on duration of exposure that is provided on the questionnaire and from the NHSRxS. Rates (incidence densities; IDs) can be calculated for a given fixed time period (t)–ID_t, for all events reported in patients for 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,

$$ID_t \text{ per 1000 patient months of treatment (}or\ observation\text{)} = \frac{N_t}{D_t} \times 1000$$

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.

IDs 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 below) 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 censoring would be different for each outcome, the denominator for the crude ID does not initially include 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 study 20.2, which describes the PEM study of drospirenone/ ethinyl estradiol.[21]

Comparisons of event rates (incidence densities)

Calculating measures of impact (ID rate differences) of 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

Case study 20.1 Finasteride and gynecomastia

Background

- Indicated for the treatment and control of benign prostatic hyperplasia (BPH)
- Finasteride is an inhibitor of 5-α reductase, which catalyzes the conversion of testosterone to dihydrotestosterone (DHT).
- Company literature at the time of marketing indicated that the most frequent adverse events were related to sexual function and that there were feminizing effects; gynecomastia was not included in the datasheet 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 PEMbase.
- Qualitative methods involved further evaluation and characterization of the event.

Results

- The PEM cohort comprised 14767 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 (Fig. 20.4).
- 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 de-challenge in 15 of the 31 men in whom 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 was 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.

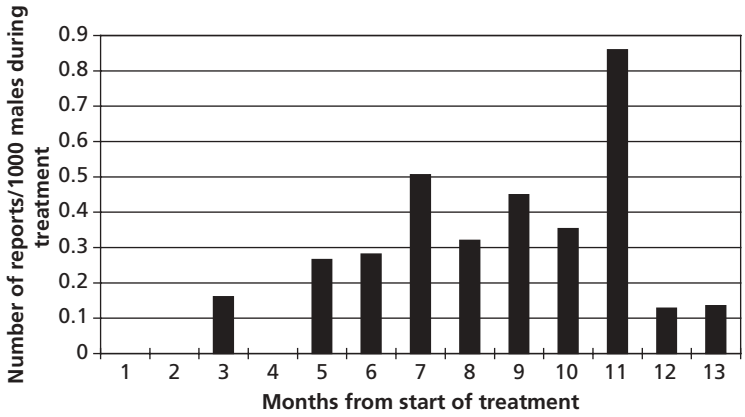


Figure 20.4 Reports of gynecomastia during treatment with finasteride.

Case study 20.2 Yasmin® and venous thromboembolism

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 can not 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 patient's 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 deep vein thrombosis and pulmonary embolism 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.
- As an association between the COC and VTE has been recognized for more than 40 years, it is important to know whether a new COC is associated with VTE and to what extent.
- The lack of selection bias and the large numbers of women studied led to identification of women who developed VTE while taking Yasmin, all of whom had risk factors of 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.

the first month after starting treatment and the IDs for months 2 to 6 (ID_1–$ID_{2–6}$) 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 (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–$ID_{2–6}$ value for an event is positive or $ID_1/ID_{2–6}$ is above 1 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 to 6 combined. This result can be considered to be 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–$ID_{2–6}$ value for an event is negative, or the ID_1/$ID_{2–6}$ is less than one and the confidence limits around the point estimate exclude the null value (zero or one respectively). In such settings, the rate of events in months 2 to 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 20.3 shows a summary of such data from a typical PEM study of a drug (oxcarbazepine)[22] for which 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/ post-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 20.4 shows a summary of such data from a PEM study of a drug (desloratadine)[23] 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 as reported in the PEM study of strontium.[24]

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 (not caused by the drug) event rate, whereas a non-constant 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 assessed. An example is the examination of hazard rates of hypoglycemia as reported in PEM studies of thiazolidinedione antidiabetic drugs.[25]

Reasons for stopping the drug

The PEM questionnaire asks the doctor to record the reason why the drug was stopped if, in fact, it was stopped. This is very informative because it includes possible adverse reactions which the doctor and/or the patient considered serious or sufficiently troublesome to stop the medication. Clinical and non-clinical reasons for stopping a drug are presented in two ways: by dictionary SOC 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 Tables 20.3 and 20.4).

Outcomes of pregnancy

All pregnancies reported during PEM 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.[26] Projects are underway to compare pregnancy outcomes following drug exposure between PEM studies or between

Table 20.3 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 $ID_1 \geq 3$)

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
Non-surgical 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
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

Events associated with starting treatment are in bold italic.
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).

Table 20.4 Incidence densities (ID per 1000 patient months) for standard PEM study for desloratadine ranked in order of ID_1 for all events during observation (between date of starting and end of observation) in the first 2 months after starting treatment (where IDobs m_1 >1)

Higher term	N_1	N_2	IDobs m_1	IDobs m_2	IDobs (m_1/m_2)	95%CI	ADR	RFS
Condition improved	***1384***	***606***	***117.04***	***51.34***	***2.28***	***2.07–2.51***	**–**	***1984***
No further request	***658***	***77***	***55.65***	***6.52***	***8.53***	***6.73–10.95***	**–**	***733***
Not effective	***537***	***238***	***45.41***	***20.16***	***2.25***	***1.93–2.63***	**–**	***772***
Course completed	***177***	***24***	***14.97***	***2.03***	***7.36***	***4.79–11.80***	**–**	***201***
Upper respiratory tract infection	53	47	4.48	3.98	1.13	0.75–1.70	0	4
Patient request	***40***	***16***	***3.38***	***1.36***	***2.50***	***1.37–4.77***	**–**	***56***
Hospital referrals no admission	***30***	***13***	***2.54***	***1.10***	***2.30***	***1.17–4.81***	**–**	***14***
Headache, migraine	***28***	***7***	***2.37***	***0.59***	***3.99***	***1.70–10.83***	***0***	***9***
Lower respiratory tract infection	24	17	2.03	1.44	1.41	0.73–2.79	0	1
Other drug substituted	23	13	1.95	1.10	1.77	0.86–3.80	–	36
Urinary tract infection	***19***	***6***	***1.61***	***0.51***	***3.16***	***1.21–9.67***	***0***	***0***
Effective	***17***	***2***	***1.44***	***0.17***	***8.49***	***2.01–75.72***	**–**	***19***
Infection skin, unspecified*/ local bacterial	17	9	1.44	0.76	1.89	0.79–4.80	0	0
Rhinitis allergic	17	18	1.44	1.52	0.94	0.46–38.93	0	5
Asthma worse	16	12	1.35	1.02	1.33	0.59–3.08	0	2
Depression	16	10	1.35	0.85	1.60	0.68–3.94	0	1
Injury	15	12	1.27	1.02	1.25	0.55–2.92	0	0
Pain abdomen	14	8	1.18	0.68	1.75	0.68–4.81	1	4
Rash	14	7	1.18	0.59	2.00	0.75–5.84	0	2
Menstrual disorder	8	9	1.13	1.27	0.89	0.30–2.59	0	0
Pregnancy	8	8	1.13	1.13	1.00	0.33–3.05	0	7
Eczema	12	10	1.01	0.85	1.20	0.47–3.10	0	1

Events associated with starting treatment are in bold italic.

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.

IDobs m_1 = Incidence density for each event during observation month 1 (where Dobs m_1 = 11825).

IDobs m_2 = Incidence density for each event during observation month 2 (where Dobs m_2 = 11804).

IDobs (m_1/m_2) = Relative difference between $IDobs_1$ and $IDobs_2$.

95% CI = 95% confidence intervals for IDobs (m_1/m_2) lower and upper.

RFS = Reasons for stopping desloratadine (Total no. reports = 3969 during months 1 and 2 of observation out of 5559 in 5502 patients (46.5% of cohort) for whole study period).

ADR = Adverse drug reactions (total no. reports = 19 during months 1 and 2 of observation out of 20 reports in 18 patients (0.15% of cohort) for whole study period.

*Unspecified: no event term currently exists in DSRU dictionary.

PEM studies and external comparators. The comparisons within the PEM database include comparing pregnancy outcomes for women who continue to take a particular drug with women who stop taking the drug. It is important that studying pregnancy outcomes continues in order to exclude, to the greatest extent possible, teratogenic effects of medicines (see Chapter 28).

Strengths

Representativeness and size

PEM uses a non-interventional exposed-only cohort design to monitor multiple outcomes in a large group of patients. It does not interfere with the prescribing decision process of the practitioner and information is collected *after* the prescribing decision has been made and implemented. This means that in PEM, 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. For example, PEM studies capture data on unlicensed and unlabelled prescribing including unlicensed prescribing for children. In addition, there are no predefined selection criteria in terms of GPs or patients based on particular characteristics. This combination makes it likely that the PEM cohort is representative of all patients who have started the study drug under similar circumstances on a national scale in England. In this way, the system avoids the problem of generalizability inherent in clinical trials, including many postmarketing safety trials.

In terms of size, the PEM database contains over 5.4 million prescriptions and data on over 1 million individuals exposed to new treatments in England. The PEM data collection method has been shown to be successful in regularly producing data on 10 000 or more patients given newly marketed drugs which, by virtue of their success in the marketplace, involve substantial patient exposure. It fulfills, therefore, the original objective of providing a prescription-based method of postmarketing surveillance of new drugs intended for widespread, long-term use.

Variable direction of investigation

PEM combines both retrospective and prospective aspects. The previous history can be studied, as well as providing opportunities for following up subgroups of patients of interest in a prospective manner. The introduction of modified PEM methods has enhanced the gathering of relevant data on exposure to risk factors, enables other factors to be considered, and enables more detailed investigations of specific safety issues of concern at the time of starting the study or those that emerge while the study is in progress.

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 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 do not get a prescription dispensed,[27] this is an advantage in that PEM exposure data are more accurate than that derived from records of physician-issued prescriptions (which are not always dispensed), as held in some pharmacoepidemiologic databases.

Outcome data

In PEM, the prescriber is asked to supply event data regardless of causality. Therefore, PEM is able to identify signals of adverse reactions or syndromes which none of the participating GPs suspect to have been due to an ADR.[4] If the GP attributes the event to the use of the drug, then the GP is asked to record this on the PEM questionnaire. In addition, the method prompts the doctor to fill in the questionnaire and does not rely on the clinician taking the initiative to report. This "prompting" effect of PEM is most important because, as described below, ADR reporting is enhanced in PEM compared to the passive Yellow Card spontaneous ADR reporting system in the UK.

As mentioned previously, the database also allows the study of diseases as well as drugs. Although bias is possible (see Weaknesses below),

event incidence estimates (which represent an unknown combination of those events occurring in the general population and those attributable to the use of the drug) are likely to be more precise than those estimated from trials because of such large patient cohort sizes. This is also in line with the early proposals of Finney on event reporting.[4]

Signal strengthening

PEM identifies patients with potential ADRs who can be studied further. More detailed information about the clinical characteristics of particular adverse events and the patient can be obtained for semiquantitative assessment of selected cases within a case series. A variety of comparisons using selections of the PEM database (drug groups, specific patient groups) can be conducted to refine signals.[28–31] Such comparisons are appropriate because the database is comprised of new drug user populations assembled at the same stage in time in the immediate postmarketing period since introduction of each product. As described previously, it is also possible to conduct external comparisons using demographic data of the population as a whole.[32] However, in contrast to PEM's data, some pharmacoepidemiologic data sources (such as the General Practice Research Database or The Health Improvement Network, see Chapter 15) have limited data on recently introduced products, which precludes reliable comparisons being made because of small sizes of the population exposed.

Participation in research

For GPs in the UK, research and academic medical practice are considered non-core activities and therefore receive no payment from the NHS. Although GPs have a duty of care to report ADRs and co-operate with requests for information from organizations monitoring public health, those GPs who participate in PEM studies do so on behalf of research and not for monetary interest—the remuneration received for completion of forms barely covers administration costs. This is also evident for other pharmacovigilance activities such as the Yellow Card spontaneous reporting scheme. Since the early 1980s, close contact has been fostered between the research staff in the DSRU and the reporting doctors. This facilitates the gathering of supplementary information on important events, pregnancies, deaths, etc. (Tables 20.1 and 20.2), 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).

Weaknesses

Single-group cohort design

As highlighted earlier, PEM is 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 multiple exposure groups, it is more efficient in terms of resources, but the principle limitation is absence of data on an unexposed comparator. Thus, 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.

Bias

Selection bias is detrimental to the validity of observational studies.[33] In PEM, it reflects the potential that the study sample is not representative of the general population. Such error cannot be adjusted for. How different a PEM cohort is compared to all other patients with the same indication receiving other health care in general practice in England cannot be assessed since, as mentioned previously, PEM does not monitor an unexposed cohort concurrently. Channeling of new drugs also introduces selection bias through preferential prescribing. Patterns of adoption of a new drug cannot be predicted, and while it may be examined in PEM, it cannot be controlled.

Non-response bias is another form of selection bias which is possible since not all PEM questionnaires are returned. The current median responses for standard PEM and M-PEM studies are 50% and 64%, respectively. In 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.

Under-reporting, including underreporting of serious and fatal adverse events, is possible in PEM since it depends on reporting by doctors. Information bias in terms of misclassification of outcome and exposure is also possible since the data depends on the accuracy and thoroughness of the GPs in diagnosis, record keeping, and reporting.

In PEM, exposure misclassification may be introduced through inaccurate calculation of exposure. 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 PEM, exposure is calculated from the date of issue of dispensed prescriptions, which means that exposure data used for PEM are more accurate than exposure data based on physician-issued prescriptions alone. Nevertheless, patients may not take all of 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.

Confounding

PEM is like all observational studies in that a major disadvantage is the inability to control for factors that might differ between groups being compared.[34] PEM is best regarded as a general safety surveillance method which generates hypotheses of safety signals of uncommon or rare outcomes. These hypothesis 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 covariates for *all* possible outcomes is limited. Thus, in 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, the adoption of modified PEM methods has provided considerable opportunities to enhance collection of supplementary data on important risk factors.

Temporality

Examinations of whether risks change over time can be affected by skewed reporting distributions or temporal changes in prescribing practice. The generalized approach to segregation of time periods in these calculations does not provide for: signals of events delayed until *after* the first month; changes in event rates as a result of important dynamics in the population at risk such as through depletion of susceptibles;[35] switchers or healthy users; and analysis of repeated events or clusters of events.

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 PEM 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 99% confidence interval has routinely been used in PEM to aid decision making, but that might be too restrictive in terms of signal generation for safety surveillance since this increases the chance of a type 2 error, that is, of missing a difference that really is there.

Setting

PEM is currently restricted to general practice. Drugs that are mainly used in hospitals cannot be studied with the current method of PEM. However, the principle method has been adapted to examine drug initiations by specialists in the secondary care setting. One such example is the Observational

Assessment of Safety in Seroquel (OASIS) Study, which is designed to examine the short-term (up to 12 weeks) safety and use of quetiapine fumarate in a prolonged-release formulation (Seroquel XL™), along with a comparator group started on the immediate-release formulation, quetiapine IR (www.dsru.org/oasis). Any patient in England will be eligible for inclusion when a clinical decision has been made to prescribe either the XL or IR preparation of quetiapine as part of normal clinical practice for schizophrenia or mania associated with bipolar disorder. This study will enable the systematic collection and reporting of safety data on patients newly initiated on treatment with quetiapine XL. Its purpose will be to provide information on a large number of such patients and the treatment they received in a mental health-care trust setting.

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 systematic review of individual case reports and aggregate data. One important example of follow-up exploration in relation to long-latency adverse events concerned visual field defects in patients receiving long-term treatment with vigabatrin.[18,36] 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 6 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 20.1 in relation to gynecomastia and finasteride.[20]

Signal strengthening and hypothesis testing

Comparison of event rates and risks

Comparisons can be used to give estimates of relative measures of associations (RR) and 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 PEM, 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 against 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 PEM 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 PEM 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 (GI) events reported in PEM study of the COX-2 selective inhibitor celecoxib in those patients with GI risk factors (past history of GI conditions, gastroirritant drugs, use of concomitant gastroprotective agents) to those without.[37] 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.[38]

Before and after studies. These 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 prior and post 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.[39] 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 risk : benefit balance of the new formulation. The matched analysis in each subset revealed that in naïve patients, dyspnea was shown to be significantly lower in the "before" reference period (RR 0.6 [95% CI: 0.40–0.88] for post- vs. pretreatment), while for switchers dyspnea was shown to be significantly higher in the "after" high risk period (RR 1.46 [95% CI: 1.02–1.81]).

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 edema, weight gain, cardiac failure, and anemia.[40] The null hypothesis was that the risk of these outcomes 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 (non-insulin-dependent) diabetes mellitus (NIDDM). Though the combination of pioglitazone and insulin is licensed and allows improvement of glycemic control, this combination is associated with increased risk of edema and may cause weight gain. The adjusted hazard ratios for each of the separate models based on PEM study data for patients taking pioglitazone–insulin combination compared to those taking pioglitazone monotherapy and/or pioglitazone with another antidiabetic (sulfonylurea or metformin) were: edema 2.28 (95% CI: 1.37–3.78); weight gain 2.03 (95% CI: 1.15–3.58), and cardiac failure 1.73 (95% CI: 0.63–4.74). This suggests that patients taking the pioglitazone–insulin combination had higher risks than pioglitazone monotherapy or pioglitazone combined with another antidiabetic drug.

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.[41] 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 PEM 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.[42]

Standardization. Where appropriate, comparisons are made between patients identified within a PEM cohort data and an external reference group, if a suitable internal reference PEM 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 a PEM 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 the PEM study was compared with external epidemiologic data for men in England.[32] The SMR for deaths reported to have been caused by ischemic heart disease in the sildenafil PEM cohort was 69.9 (95% CI: 42.7–108.0). Although the point estimate suggests that there are 30.1% fewer deaths, the CI includes the null (1.0) which means that one cannot be certain that the death rates truly differ between the sildenafil PEM cohort and the general population reference cohort. 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.[43] Obviously, there is higher potential for bias when using external comparators than comparisons undertaken between PEM 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 is exploring the use of data-mining disproportionality methods that are commonly used in pharmacovigilance (see Chapter 10) as a possible additional quantitative tool in PEM for signal generation, because of the large number of drug–event combinations held in the PEM database. Feasibility studies have employed proportional reporting ratios (PRRs)[44] 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.[45,46] 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. The DSRU has also piloted this extension as a tool to detect patterns of adverse events associated with pharmacologic aspects such as biochemical aspects and/or structure of a drug class. One example was to model the relationship between risks of serious GI and thrombotic vascular events relative to the drug-specific COX-2/COX-1 selectivity ratio.[47]

There are a number of methodologic issues that need to be further examined 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, that is higher- or lower-level terms. However, with refinement, automated signal generation is likely to prove a useful tool to support signal generation for PEM through other quantitative methods as described above.

Epidemiology of diseases and indications

The PEM database allows the study of diseases as well as drugs. An example includes a study using data from 58 completed PEM studies carried out in the period between September 1984 and June 1996 of the prevalence of Churg–Strauss syndrome and related conditions in patients with asthma.[48] The study defined the study period prevalence rate for this condition, 6.8 (95% CI: 1.8–17.3) per million patient-years of observation and demonstrated a much higher period prevalence rate in patients receiving asthma medications (nedocromil, salmeterol, and bambuterol) of 64.4 per million patient-years of observation compared to 1.8 per million patient-years of observation in the 55 other drug cohorts. In another study, the PEM database was used to quantify age- and gender-specific asthma death rates in patients using long-acting beta-2 agonists.[49]

Drug utilization

Drug utilization research (see Chapter 24) is an essential part of pharmacoepidemiology, as it describes the extent, nature, and determinants of drug exposure at the patient level. Data from PEM 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 ongoing M-PEM study of ivabradine (which is licensed for chronic stable angina) and its utilization in patients under 40 years of age, which is likely to be used for other indications since angina prevalence is expected to be low in this group.[50] 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 GI conditions (i.e., were at low risk).[37,51]These observations were discordant with national NSAID prescribing practice during the time these drugs were first marketed, and agree with findings from elsewhere.[52]

Predictors of risk

The nested case–control design is particularly advantageous for studies of predictors of disease. The method overcomes some of the disadvantages associated with non-nested case–control studies while incorporating some of the advantages of a cohort study (see Chapter 47).[53] 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. PEM 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. Two prospectively designed nested case–control studies are underway to investigate the association between dose and the occurrence of two outcomes (extrapyramidal symptoms; somnolence and sedation) in users of a new formulation of an atypical antipsychotic.

Monitoring drug safety in children

The safety of medicine 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 will have to be included in the benefit–risk management plan (see Chapter 29) and other pharmacovigilance activities. Therefore, pharmacovigilance tools may need to be adapted to examine specific issues associated with this special population. Given that PEM 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 the PEM database. 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 incidence densities (IDs) examined between two time periods after starting treatment (month 1 and months 2–6 combined). Proportional reporting ratios 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.

Quantifying ADR reporting

Study of the PEM database also shows some of the characteristics of ADR reporting. Two studies, conducted on different PEM studies and over different time periods, compared events that were considered as ADRs by doctors reported in PEM with spontaneous reports sent by the same doctors to the regulatory authority.[55,56] 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.[55] 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.[56] This represents an under-reporting rate of 91% in both studies. It is of interest that a higher proportion of serious than non-serious reactions were reported to the CSM 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 worry them most.

It is possible to use PEM 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.[57] 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.

Assessment of therapeutic risk management programs

Therapeutic risk management is attracting immense interest in pharmacovigilance (see Chapter 29). The management of risk of medicines requires identification, measurement, and assessment of risk, followed by risk–benefit evaluation, then taking actions to eliminate or reduce the risk, followed by methods to monitor that the actions taken achieve their objectives. PEM contributes not only 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 that has recently been completed is that of the antidiabetic agent, pioglitazone.[58] Undertaken after the completion of the PEM study, detailed questionnaires were sent to doctors who reported selected adverse events such as liver function abnormalities or fluid retention to study how these events were detected and managed by doctors, as well as their outcomes. The objective was to assess the compliance at the patient level with the risk management requirements for these products. Another study was conducted to monitor the introduction of carvedilol for the treatment of cardiac failure.[59] The product (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 is being managed in the community, for example what 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 2 years.

Since risk management of medicines became a regulatory requirement in Europe in 2005, a number of modified PEM studies have been undertaken to address specific questions related to detailed examination of particular adverse events and studying drug utilization patterns (Table 20.5). 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 developing a risk management plan for the evaluation of the safety a new medicine. The DSRU and the PEM method are registered within the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Database of Research Resources (see Chapter 6), which serves as a central resource for both researchers and study sponsors seeking to identify organizations and data sets for

Table 20.5 Design and applications of M-PEM 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*:[59] 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*:[60] 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 ®)*:[61,62] EMA produced guidelines for the conduct of postmarketing surveillance studies to assess the introduction of CFC-free inhalers. M-PEMs collected data 3 months pre- and postexposure 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*: 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*:[50] 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 is investigating 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*: 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 includes prospective nested matched case control study to explore relationship between dose and events of somnolence and EPS.

conducting specific pharmacoepidemiologic and pharmacovigilance studies in Europe.

The future

In the future, PEM aims to utilize improvements in information technology. Plans are underway to expand the application of additional study designs such as nested case–control studies or self-controlled case series analysis and the application of new biological developments such as pharmacogenetics to enhance the PEM process (see below). Modification of the PEM method will continue to evolve to examine specific drug safety questions, for targeted outcome surveillance related to risk management of marketed medicinal products.

Self-controlled case series analysis

Other methodologic developments that are being introduced to M-PEM studies to examine temporal associations between specific events of interest and starting treatment with a new drug include the application of the methods of self-controlled case series studies proposed by Farrington.[63] The method was originally developed to study adverse reactions to vaccines. The method uses only cases, no separate controls are required as the cases act as their own controls, thus minimizing the effect of confounding by factors that do not vary with time, such as genetics and gender. Each case's given observation time is divided into control and risk periods. Time-varying confounding factors such as age can be allowed for by dividing up the observation period further into age categories. Because the method requires time-varying covariate data on cases only and not for the whole cohort, it is efficient in terms of sample size and resource. The method requires that specific criteria are met (for example occurrence of the event of interest should not affect subsequent exposure history or increase mortality). Using this approach, relative risk estimates are automatically adjusted for all fixed confounders. Non-cases can be ignored without bias while cases are self-matched. Conditional regression modeling will provide the adjusted estimate of relative incidence (with 95% confidence intervals) of the outcome for the high risk observation period of interest relative to the remaining observation time. PEM studies provide an ideal platform to enable the relative incidence of newly diagnosed outcomes of interest to be studied between predefined high and low risk periods in new users, thus enabling time-to-occurrence of selected events to be explored and reviewed for evidence of temporal patterns.

Pharmacogenetics

There is increasing interest in understanding the role pharmacogenetics plays in the efficacy and safety of medicines (see Chapter 34). The main purpose of pharmacogenetics is to guide effective pharmacologic treatments while minimizing ADRs. Given the interest in understanding the influence of inherited variation in drug metabolizing enzymes, receptors, and drug transporters, there are many opportunities in PEM to study the individual genetic profile of patients who develop selected ADRs compared to patients who do not develop such ADRs. Detection of specific genetic biomarkers also allows the identification of patients who do not respond to some medications. Due to the ease of accessibility to genetic information through peripheral blood or saliva sampling, as well as advances in molecular techniques, there is scope for incorporation of pharmacogenetic testing within large PEM cohort studies. In addition, the nested-case–control study design can be used to detect differences in how individual patients respond to certain medications. Pharmacogenetics within PEM has the potential to minimize ADRs and maximize therapeutic benefit for individual patients.

Conclusion

PEM contributes to the better understanding of the safety of medicines. Both signals generated by PEM and those generated in other systems and studied further by PEM have been useful to inform the debates on the safety of medicines, including supporting public health and regulatory decisions. In addition, the breadth of the PEM database provides opportunities for research on disease epidemiology

and risk management of adverse drug reactions. Like all scientific approaches, PEM is evolving, aiming to reduce its weaknesses and enhance its strengths. The most significant development of PEM in the last few years has been the introduction of M-PEM studies that obtain more information about background history and baseline details of clinical information as well as more details about specific events. M-PEM also provides opportunities for comparisons of event rates of different drugs. New methodologic modifications and additions include more effective utilization of information technology and statistics, as well as the application of new study designs such as nested case–control and pharmacogenetic studies. Pharmacovigilance and pharmacoepidemiology are emerging and exciting disciplines with evolving study methods. PEM continues to contribute to the progress of these important scientific and public health disciplines.

Acknowledgment from the director

PEM 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 NHSRxS; PEM 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 material from the previous editions and to Georgina Spragg who helped in locating research material.

References

1. Lenz W, Knapp K. Thalidomide embryopathy. *Arch Environ Health* 1962; **5**: 100–5.
2. Committee on Safety of Drugs. *Report for 1969 and 1970*. London: HMSO, 1971.
3. Sachs RM, Bortnichak EA. An evaluation of spontaneous adverse drug reaction monitoring systems. *Am J Med* 1986; **81**(5B): 49–55.
4. Finney DJ. The design and logic of a monitor of drug use. *J Chron Dis* 1965; **18**: 77–98.
5. Wright P. Untoward effects associated with practolol administration: oculomucocutaneous syndrome. *Br Med J* 1975; **1**(5958): 595–8.
6. Inman WHW. Postmarketing surveillance of adverse drug reactions in general practice. I. Search for new methods. *Br Med J* 1981; **282**: 1131–2.
7. Inman WHW. Postmarketing surveillance of adverse drug reactions in general practice. II: Prescription-event monitoring at the University of Southampton. *Br Med J* 1981; **282**: 1216.
8. Council for International Organizations of Medical Sciences (CIOMS), World Health Organization (WHO). *International Ethical Guidelines for Biomedical Research Involving Human Subjects*, 2002. http://www.cioms.ch/publications/layout_guide2002.pdf (accessed July 2011).
9. Royal College of Physicians of London. *Guidelines on the Practice of Ethical Committees in Medical Research involving Human Subjects*. London: RCP, 1996.
10. Department of Health. *Supplementary Operational Guidelines for NHS Research Ethics Committees, November 2000. Multi-centre Research in the NHS the Process of Ethical Review when there is no Local Researcher*. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4014561.pdf (accessed July 2011).
11. General Medical Council. *Confidentiality*. http://www.gmc-uk.org/static/documents/content/Confidentiality_0910.pdf (accessed July 2011).
12. Epstein M. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2005; **14**: 589–95.
13. National Information Governance Board. Register of Approved Section 251 Applications. http://www.nigb.nhs.uk/s251/registerapp (accessed July 2011).
14. European Medicines Agency. ICH Topic E2D Post Approval Safety Data Management. Step 5 .Note for guidance on definitions and standards for expedited reporting. 2004 (CPMP/ICH/3945/03). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002807.pdf (accessed July 2011).
15. Strom B. Sample size considerations for pharmacoepidemiology studies. In: Strom B, ed. *Pharmacoepidemiology*, 2nd edn. Chichester, UK: John Wiley & Sons, 1994.
16. Machin D, Campbell M, Fayers P, Pinol A. *Sample Size Tables for Clinical Studies*, Table 7.1. Oxford, UK: Blackwell Science, 1997, p. 150.

17. Machin D, Campbell M, Fayers P, Pinol A. *Sample Size Tables for Clinical Studies*, Table 7.2. Oxford, UK: Blackwell Science, 1997, pp. 151–3.
18. Wilton LV, Stephens MD, Mann RD. Visual field defect associated with vigabatrin: observational cohort study. *BMJ* 1999; **319**: 1165–6.
19. Layton D, Marshall V, Boshier A, Friedmann P, Shakir SA. Serious skin reactions and selective COX-2 inhibitors : a case series from prescription-event monitoring in England. *Drug Saf* 2006; **29**: 687–96.
20. 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**: 379–84.
21. Pearce HM, Layton D, Wilton LV, Shakir SA. Deep vein thrombosis and pulmonary embolism reported in the Prescription Event Monitoring Study of Yasmin. *Br J Clin Pharmacol* 2005; **60**: 98–102.
22. Buggy Y, Layton D, Fogg C, Shakir SA. Safety profile of oxcarbazepine: Results from a prescription-event monitoring study. *Epilepsia* 2010; **51**: 818–29.
23. Layton D, Wilton L, Shakir SA. Examining the tolerability of the non-sedating antihistamine desloratadine: a prescription-event monitoring study in England. *Drug Saf* 2009; **32**: 169–79.
24. Osborne V, Layton D, Perrio M, Wilton L, Shakir SA. 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**: 579–91.
25. Vlckova V, Cornelius V, Kasliwal R, Wilton L, Shakir SA. Hypoglycaemia with oral antidiabetic drugs: results from prescription-event monitoring cohorts of rosiglitazone, pioglitazone, nateglinide and repaglinide. *Drug Saf* 2009; **32**: 409–18.
26. 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**: 882–9.
27. Beardon PH, McGilchrist MM, McKendrick AD, McDevitt DG, MacDonald TM. Primary non-compliance with prescribed medication in primary care. *BMJ* 1993; **307**: 846–8.
28. Layton D, Clark DW, Pearce GL, Shakir SA. Is there an association between selective serotonin reuptake inhibitors and risk of abnormal bleeding? Results from a cohort study based on prescription event monitoring in England. *Eur J Clin Pharmacol* 2001; **57**: 167–76.
29. Layton D, Heeley E, Shakir SA. 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**: 171–81.
30. Layton D, Wilton L, Boshier A, Cornelius V, Harris S, Shakir SA. Comparison of the risk of drowsiness and sedation between levocetirizine and desloratadine : a prescription-event monitoring study in England. *Drug Saf* 2006; **29**: 897–909.
31. Acharya NV, Pickering RM, Wilton LW, Shakir SA. The safety and effectiveness of newer antiepileptics: a comparative postmarketing cohort study. *J Clin Pharmacol* 2005; **45**: 385–93.
32. Boshier A, Wilton LV, Shakir SA. Evaluation of the safety of sildenafil for male erectile dysfunction: experience gained in general practice use in England in 1999. *BJU Int* 2004; **93**: 796–801.
33. Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health* 2004; **58**: 635–41.
34. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 2003; **12**: 551–8.
35. Moride Y, Abenhaim L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol* 1994; **47**: 731–7.
36. Stephens MDB, Wilton LV, Pearce G, Mann RD. Visual field defects in patients taking vigabatrin. *Pharmacoepidemiol Drug Saf* 1997; **6** (Suppl. 2): S18.
37. Layton D, Wilton LV, Shakir SA. Safety profile of celecoxib as used in general practice in England: results of a prescription-event monitoring study. *Eur J Clin Pharmacol* 2004; **60**: 489–501.
38. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med* 1991; **10**: 577–81.
39. Osborne V, Layton D, Tong E, Shakir SA. Safety of Atrovent CFC-free MDI (metered dose inhaler): An example of protopathic bias in observational cohort studies (oral presentation). *27th International Conference on Pharmacoepidemiology and Therapeutic Risk Management*, Chicago, Aug. 2011. In press.
40. Kasliwal R, Cornelius V, Wilton LW, Shakir SAW. Abstract no: 255. Combined Pioglitazone-Insulin Use: Analysis from the Pioglitazone PEM Study. 24th International Conference on Pharmacoepidemilogy and Therapeutic Risk Management. Copenhagen, Denmark, August 17–20, 2008. *Pharmacoepidemiol Drug Saf* 2008; **17** (Suppl 1): S1–294.

41. Layton D, Heeley E, Hughes K, Shakir SA. 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* (Oxford) 2003; **42**: 1342–53.
42. Layton D, Harris S, Wilton LV, Shakir SA. 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**: 473–82.
43. Boshier A, Wilton LV, Shakir SA. 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**: 767–73.
44. 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**: 483–6.
45. Mackay FJ, Wilton LV, Pearce GL, Freemantle SN, Mann RD. Safety of long-term lamotrigine in epilepsy. *Epilepsia* 1997; **38**: 881–6.
46. Heeley E, Wilton LV, Shakir SA. Automated signal generation in prescription-event monitoring. *Drug Saf* 2002; **25**: 423–32.
47. Layton D, Ferreira GLC, Shakir SAW. Abstract no: 305. Integrating pharmacology with pharmacovigilance: developing tools to examine pattern of risks with drug classes in prescription-event monitoring (PEM). 23rd International Conference on Pharmacoepidemiology and Therapuetic Risk management (ISPE), 19–22th August, 2007, Quebec, Canada. *Pharmacoepidemiol Drug Saf* 2007; **16** (Suppl S2): S1–S274.
48. Martin RM, Wilton LV, Mann RD. Prevalence of Churg–Strauss syndrome, vasculitis, eosinophilia and associated conditions: Retrospective analysis of 58 prescription-event monitoring cohort studies. *Pharmacoepidemiol Drug Saf* 1999; **8**: 179–89.
49. Martin RM, Dunn NR, Freemantle SN, Mann RD. Risk of non-fatal cardiac failure and ischaemic heart disease with long acting beta 2 agonists. *Thorax* 1998; **53**: 558–62.
50. Doe C, Fogg C, Layton D, Shakir SAW. Abstract no: 330. Interim results of a modified prescription event monitoring study on ivabradine: case series of utilisation in patients under the age of 40 years, 26th International Conference on Pharmacoepidemilogy and Therapeutic Risk Management. Brighton, England, August 19–22, 2010. *Pharmacoepidemiol Drug Saf* 2010; **19** (Suppl 1): S1–S347.
51. Layton D, Riley J, Wilton LW, Shakir S. Safety profile of rofecoxib as used in general practice in England: results of a Prescription-Event monitoring study. *Br J Clin Pharmacol* 2002; **55**: 166–74.
52. Layton D, Souverein PC, Heerdink ER, Shakir SA, Egberts AC. 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**: 143–58.
53. Flanders WD, Louv WC. The exposure odds ratio in nested case-control studies with competing risks. *Am J Epidemiol* 1986; **124**: 684–92.
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**: 751–63.
55. 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**: 119–20.
56. Heeley E, Riley J, Layton D, Wilton LV, Shakir SA. Prescription-event monitoring and reporting of adverse drug reactions. *Lancet* 2001; **358**: 1872–3.
57. 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**: 505–11.
58. Fogg C, Kasliwal R, Shakir SA. Risk management and outcomes of adverse events to pioglitazone in primary care in the UK: an observational study. *Drug Saf* 2009; **32**: 229–37.
59. Aurich-Barrera B, Wilton LV, Shakir SA. 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**: 43–54.
60. Davies MN, Paiba N, Wilton LW, Shakir SAW. Abstract no: P.020. A 12 month modified prescription-event monitoring report for travoprost. 7th Annual conference of ISOP, 21–24, October, 2007. Bournemouth, England. *Drug Saf* 2007; **30**: 919–90.
61. Perrio MJ, Wilton LV, Shakir SA. 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**: 969–78.

62. Perrio MJ, Wilton LV, Shakir SA. 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**: 681–95.
63. Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol* 1996; **143**: 1165–73.

CHAPTER 21
Registries

Nancy A. Dreyer and Priscilla Velentgas
Outcome Sciences Inc., Cambridge, MA, USA

Introduction

Registries are special purpose or sometimes broad observational data collection programs that can provide a natural context for understanding effectiveness, safety, and patterns of care in defined populations. In this context, effectiveness may be described as the extent to which medical interventions achieve health improvements in real practice settings. Registries can be thought of as programs that collect data from which studies may be derived; the term "registry" may be also be used to describe the actual database.

The concept of registering and counting people because of a shared characteristic has a long tradition, with references in biblical times to registries of births (Psalm 87:6) and population registries (2 Samuel 24:4). In public health, registries of burials and christenings were used to draw inferences about the onset and spread of bubonic plague.[1] More recently, epidemiologists use registries as data sources for studies of the effectiveness and safety of a wide variety of medical interventions. These registry-based studies are sometimes referred to simply as "registries," "registry-based studies," or by their study design, for example cohort or case–control studies.

Traditionally, registries were either population-based tools for monitoring public health interventions, such as registries that recorded receipt of childhood vaccines, or collections of data on people with shared characteristics, such as disease registries or other systematic programs for case ascertainment and recruitment. Recently however, registry methods have been formalized, supporting the transformation of many programs from narrow purposes, such as patient support activities, to scientific research programs that can be used to evaluate effectiveness and safety of marketed products and other health interventions (see Chapter 3), contribute information that can be used for regulatory and payment related decisions, and as support for controlled distribution programs and risk evaluation and mitigation strategies (REMS, see Chapter 29). Registries are also widely used for characterizing diseases in terms of clinical presentation and progression. Some examples of the specific and more open-ended types of questions a registry can answer include:

- What is the rate of a specific adverse event among users of a particular drug (such as the rate of suicide among users of isotretinoin)?
- What are the long-term health outcomes in children who are treated with tumor necrosis factor (TNF)-inhibitors for rheumatic disease?
- Prior to the availability of enzyme replacement therapy, what was the natural history of Gaucher's disease, a lysosomal storage disorder?

While clinical trials make strong contributions to pharmacoepidemiology and particularly to comparative effectiveness (see Chapters 32 and 36), clinical trials have known limitations due to the type of population studied (optimal patients and medical care providers), the way the interventions

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

are administered and analyzed (controlled dosing, limited use of other medications, often in unusual settings with highly experienced practitioners, etc.), the lack of real-world comparator data, and the often short follow up. Thus, there are ample opportunities for well-designed and well-conducted observational studies, whether using registries or not, to help fill the evidence gap by complementing the highly controlled settings of premarket and postmarket clinical trials in addressing questions about the safety and effectiveness of pharmaceutical, biological, and device products, and medical interventions that are important to public health.[2–5]

How do registries differ from traditional cohort studies? In its purest sense, the term "registry" applies to the database from which a cohort could be constructed. Since some registry-based studies are analyzed using cohort methods and are considered to be cohort studies, distinctions between the two have more to do with the objectives of the research program than aspects of study design. In common practice, the designation of a "registry" generally refers to an ongoing research program, for example a program that records all seasonal influenza vaccines or a study of a given condition, treatment, or procedure over time. Registries often collect detailed clinical information at the point of patient care, such as laboratory values and physician or patient assessments of disease severity and quality of life. This approach is also often used to address several research questions, and the purpose and scope of a registry often adapts over time. For example, registries that are designed to evaluate product safety also are often used to evaluate effectiveness, to estimate the size of various patient subgroups of interest, to characterize physicians who treat patients of interest, etc. They might also provide a forum for patients to share information and support, and for physicians to learn from each other about best treatment practices. Registries often start with a specific focus and collect data for that purpose; then as knowledge develops, new research questions surface and the data collection is either modified or supplemented. Thus, a common trait of registries is their flexible approach to research.

Description

Registry design

Generally the term "registry" is applied to programs in which patients are sampled from discrete, well-defined populations, or when typical patients and physicians are recruited for study (see Box 21.1). Each enrollment strategy has its benefits, including practical feasibility, and restrictions on interpretation. For example, a registry derived from a well-defined underlying population can be used to estimate incidence and prevalence of the condition and to describe natural history. A registry of typical patients can be used to characterize treatment effectiveness and safety, and to estimate the incidence of adverse treatment effects, etc., but not to estimate incidence or prevalence of a condition in a population, since the baseline population has not been enumerated. It is critical to understand how a registry is constructed, and in particular, how patients come to be included in the study, in order

Box 21.1 Definition of a patient registry

A *registry* is an organized system that uses 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.[6] The unit of observation may be the patients or the treatment of interest. For example, in some registries, multiple products may be used by the same patient, such as artificial joints or implanted heart valves.

Registries may be drawn from a well-defined underlying population such as a health care system or a geographical area (region, state, or country), may reflect 100% of the treated population as in a controlled distribution program, or may represent people who enroll on a voluntary basis, with inclusion and exclusion criteria designed to recruit patients typical of those with the exposure or disease of interest. Registries may have internal comparators or may focus exclusively on a single intervention or method of health care delivery. The inferences that may be drawn from a registry-based study differ depending on how the study population is sampled and what is studied directly.

Depending on the scope of the registry, it can then be used as the source population for cohort studies or case–control studies.

to understand what inferences can be drawn from a registry-based study and what potential sources of bias to be alert for in the design and analysis.

Registries may have internal comparators or may focus exclusively on a single intervention or method of health care delivery. Consider a disease registry that collects information on all patients diagnosed with a given condition of interest. Comparisons can be made among different treatments and between untreated and treated groups. Internal comparators give the advantage of providing contemporaneous information that can be used to evaluate whether the observed effects of a given treatment are meaningfully different from those of another treatment observed in a similar population during the same time period using equivalent data collection methods. When internal comparators are not included, external comparators from the same time period or historical controls can be used to give context to registry-derived data. Selecting comparators for registries is always challenging because, like all non-randomized studies, treatment assignment is not random and selection factors such as channeling bias and temporal changes need to be considered.

Some registries collect information only at a single point in time and are analyzed as cross-sectional studies, such as vaccine registries.[7–9] While these registries can be important tools for public health, there is greater interest and emphasis for pharmacoepidemiology in registries that permit follow up over time. Registries with follow up can be used for cohort studies and can extend for weeks, months, or even decades. The National Registry for Myocardial Infarction[10] followed patients through the duration of a single hospital admission for myocardial infarction, whereas the Cystic Fibrosis Registry[11] follows young children with cystic fibrosis through adulthood. Many registries adapt over time to include new research objectives and additional data collection as information accumulates and needs change. For example, a global disease registry focused on understanding treatment effectiveness for avian influenza has adapted over time from an initial focus on antivirals to collecting more clinical variables to promote understanding differences between pediatric and adult populations, and the prognostic value of various clinical signs and symptoms as patients present for medical attention.[12,13]

Like any research study, size matters in terms of the likelihood of being able to detect meaningful effects should any exist (see Chapter 4). Larger sample sizes also provide the potential for subgroup analyses of special populations at risk. Longer follow up allows for evaluation of delayed benefits and risks. Even registries with few patients or relatively little accrual of person-time, if well designed, can potentially be very useful, particularly when there is little other evidence. For example, a registry was used to study a rare form of vaginal cancer in women exposed to diethylstilbestrol *in utero*.[14] The registry of clear-cell adenocarcinoma of the lower genital tract in young women became an important long-term research tool for understanding risks to female offspring[15] and later was extended to study health effects in male offspring.

Taxonomy

A common "taxonomy" of registries classifies them according to the characteristics of the population that are the focus of study (Table 21.1).[6] These include the following groupings (see page 334):

Table 21.1 Registry taxonomy with examples

Type of Registry	Example (sponsor)
Disease or event registry	International Collaborative Gaucher Group Gaucher Registry[16] (Genzyme) National LymphoCare Study[17] (Genentech and Biogen Idec)
Product registry	Implantable cardiac device registry (Agency for Healthcare Research and Quality, American College of Cardiology) iPledge Isotretinoin Registry (all manufacturers of generic isotretinoin)
Service or procedure registry	SAPPHIRE registry for carotid artery stenting[18] Get With the Guidelines–Coronary Artery Disease[19] (American Heart Association)

• *Disease or event registries:* inclusion of subjects based on diagnosis of a common disease or condition
• *Product registries:* inclusion of subjects based on use of a specific product (drug or device) or related products in a given therapeutic area; pregnancy registries are a subcategory
• *Service or procedure registries:* inclusion of subjects based on receipt of specific services, such as procedures, or based on hospitalizations.

Data sources

Registry data may be drawn from a number of sources. Though often involving prospective data collection from treating physicians, registries may incorporate any of the following sources of data:
• Clinical and medical data collected from healthcare providers and/or patients, including patient-reported outcomes
• Paper or electronic health records
• Electronic administrative and billing data.

Registries with limited objectives may be populated from secondary data sources such as routinely collected electronic data. For example, administrative claims data are used to populate US state immunization registries such as the I-CARE (Illinois Comprehensive Automated Immunization Registry Exchange) developed by the Illinois Department of Health to facilitate information sharing among health care providers.[20] However, it is common that some data are collected *de novo,* specifically for the purposes of the registry. Ideally, enrollment of patients in a registry is a means of leveraging multiple data sources to characterize patients' medical history and disease state, treatments, and health and utilization outcomes in detail, allowing many questions of interest to be addressed. For example, linkage of registry data with other data sources, including vital records and billing data, may allow additional outcomes to be evaluated, such as clinical outcomes, including mortality, quality of life, absenteeism, burden of illness, and cost.

Quality guidelines applicable to registries

Registry-based studies may be conducted under the same general guidance used for observational studies, such as the Guidelines for Good Pharmacoepidemiologic Practice developed by the International Society for Pharmacoepidemiology,[21] and according to the same epidemiologic methods and principles applicable to cross-sectional, cohort, and nested-case control, or case–cohort studies (see Chapter 3).[22] Guidelines such as STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) for reporting of observational study results,[23] and the GRACE (Good Research for Comparative Effectiveness) Principles[24] for conducting and evaluating observational studies of comparative effectiveness, are also applicable to registry-based studies.

However, registry guidance generally needs to be tempered by a good understanding of the practical realities of studying medical care and use of medications in real-world settings. The US Agency for Healthcare Research and Quality (AHRQ) commissioned a guide to patient registries, *Registries for Evaluating Patient Outcomes*.[6] (Individual copies may be obtained at no charge at http://effectivehealthcare.ahrq.gov). Contributors from academia, government, and industry provided content that includes many examples from the US and abroad, with a focus on designing and implementing registries to provide accurate and reliable evidence to inform health care decision making. Topics covered include registry planning and design, use of registries in product safety assessment and for adverse event reporting, data sources and data linkage, and analysis and interpretation, with recent case studies.

Strengths

Patient registries can greatly facilitate research. Their ability to collect information on many factors simultaneously allows study of a broad variety of exposures, potential confounders, effect modifiers, and clinical outcomes. Detailed information can be obtained about disease characteristics, treatments, and outcomes, such as disease severity and flares, medication sequencing, and treatment combinations including pharmacological, surgical, and other interventions that might influence the course

of disease. Collection of detailed and varied exposure data prospectively is a strong distinguishing factor for registry-based studies compared with studies that use routinely collected health information such as administrative claims or medical records. Registries can be used to collect more detailed information such as product lots, batch numbers, non-prescription drugs and non-covered drugs, genetic information, and information about how medications are actually used (e.g., pill splitting, adherence, etc.), how treatments are administered (e.g., variations according to physician specialty), and why treatment decisions are made.

Another strength is the ability to adapt over time to accommodate new research questions and purposes, and to adjust processes and procedures if it becomes apparent that the initial approach is no longer feasible, appropriate, or as informative as desired. For example, a registry may need to be refocused or adapted midcourse because of the availability of a new treatment or the withdrawal of a key comparator. Periodic analyses may lead to new hypotheses that require collection of supplementary information for further evaluation.

Registry-based studies and other observational studies can provide results that more accurately reflect patient outcomes than do clinical trials because they are generally designed to be inclusive and broad with regard to both patient entry criteria and site selection, better reflecting the effectiveness and safety of medical interventions in real-world practice. For example, observational studies of surgical procedures often reveal very different results than experimental data collected from studying highly skilled surgeons in tightly controlled settings. A 1998 multivariate analysis of Medicare data showed that 30-day mortality following carotid endarterectomy (CEA) for Medicare patients in 1992 and 1993 was 1.75% overall, and was 43% lower among 89 hospitals that had participated in one of two clinical trials of CEA (the North American Symptomatic Carotid Endarterectomy Trial [NASCET] and the Asymptomatic Carotid Atherosclerosis Study [ACAS]) than low-volume non-trial hospitals.[25] The 30-day mortality in the Medicare analysis was also notably higher for the trial hospitals (1.4%) than that reported from studies of other patients in clinical trials (0.6% for NASCET, and 0.1% for ACAS). These results illustrate the impact of both clinical site selection factors and patient selection factors on conclusions that may be drawn regarding patient outcomes following health interventions.

Registry studies also provide the opportunity to follow patients over long periods of time for a variety of outcomes, as is true of other types of prospective cohort studies. For long-term chronic diseases, patients may be followed for years by their treating physicians. 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. For example, the VIRGO study follows patients with metastatic breast cancer to understand the impact of treatment on quality of life as the disease progresses.[26] It is often the patient's relationship with the physician, and the physician's relationship with the registry, that make registries particularly effective for long-term retention.

Weaknesses

The multipurpose nature of registries sometimes means that registries created to learn about a disease, treatment, or health care delivery system are often organized to answer broad questions. Registries that are designed primarily to address descriptive aims or to provide an adaptive framework for evaluating new questions about treatment of a particular condition may lack a focused hypothesis. Some consider the absence of prespecified hypotheses as diminishing the usefulness of results from registries, especially in the context of comparative effectiveness. The idea that findings identified through exploratory data analysis are unscientific comes in part from the science of clinical trials, where studies are designed around a single, prespecified main hypothesis, and serendipitous findings are not considered to be valid, no matter how strong or clinically meaningful. In

contrast to testing, the science of discovery and explanation is based on informed analyses, but not restricted only to ideas identified in advance.[2] In fact, most medical discoveries occur through explanatory analyses. Nonetheless, it is important to recognize, especially in the context of comparative effectiveness, that many outcomes researchers and others hold strongly to the position of the importance of prespecified hypotheses, and the validity of findings from registry-based studies that were not created to address a prespecified hypothesis may be considered suspect.

Registries that only include patients treated with a medical intervention of interest may be of limited use because of the difficulty of explaining whether the observed effects are due to the intervention under study or are merely a characteristic of the type of people under study. For example, in a registry-based study of men receiving a particular medical treatment for baldness, one might attribute the high rate of cardiovascular events to treatment unless comparator data were available, which would reveal this higher risk is common among bald men and is not unique to a treatment. While the inclusion of contemporaneous, internal comparators provides tremendous advantages, registries, like other research tools, are often designed with budgetary or regulatory restrictions which make it undesirable or impractical to include patients undergoing a variety of treatments. In these situations, external data can be used to aid in the interpretation of registry-based study data, and the challenges then relate to practical issues of different data, coding, groupings, etc., access to raw data or reliance on published tabular data, and to interpretation. External comparison data generally come from different populations, different geographical regions, and different time periods. Although some of the differences between groups may be known and measurable, these differences increase the risk of unknown confounding in comparison with registry data.

Another weakness of registries relates to selection and recruitment of subjects. Depending on the approach used to identify participating sites or patients, the underlying sampling frame may not be readily known and cannot be assumed to represent a random sample of a given population. Instead subjects frequently are selected because they come to medical attention through the particular recruitment scheme, for example, a particular health system. The potential for broadly generalizable registry data can be enhanced, in part, through recruitment plans. Some registries include many types of health systems and require sequential enrollment of eligible patients so that physicians cannot preferentially recruit optimal patients. Sequential enrollment can be documented using enrollment logs, which are auditable. The characteristics of the enrolled study population can also be compared with institutional data or publications on similar patients to describe how study patients compare to the larger population of interest. Like any observational study that does not systematically sample from a well-defined underlying sampling frame or which relies on voluntary participation and informed consent, the characteristics of the enrolled study population should be considered in interpreting the results.

Other operational challenges relate to data collection systems that are easy to use (e.g., multiple methods of data entry such as internet, interactive voice, fax, and/or mail), have interoperability with electronic medical records or claims data to avoid duplicate data entry, and are simple enough to encourage steady reporting. Retention rates are also higher when registries are responsive to the needs of patients and physicians so that they are motivated to continue participating (a special concern for pregnancy registries and other vulnerable populations). 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, registries that include 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. It is important to attend to all regulatory and ethical concerns relating to confidentiality, privacy, and data security. For more information, see the AHRQ User's Guide for Patient Registries.[6]

Particular applications

Registries are used to describe special conditions and to characterize the natural history of a disease and patterns of treatment and health care utilization. They are also frequently used for safety and other postmarketing product surveillance needs, including risk evaluation and mitigation strategies (REMS) programs and restricted distribution programs (see Chapter 29). Some specific applications are described further in this section.

Disease and condition registries

Registries have a longstanding record of being used to study a wide variety of diseases and conditions, including both common and rare diseases. Disease registries have been funded by government, industry, physician organizations, and philanthropic groups. For example, the US National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), supports a number of registries of rare and more common medical conditions, including lupus and neonatal lupus, to provide a data platform for studying the genetics, treatment, and other features of these diseases.[27] A private charitable foundation (Lupus Clinical Trials Consortium, a charitable foundation dedicated to the facilitation of scientific research regarding the disease of lupus) is conducting a long-term registry to describe the natural history of systemic lupus erythematosus and the clinical course of diagnosis and treatment. Some well-known special applications are described below.

Cancer registries

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 has provided statistics for monitoring of cancer disease burden since 1973, drawing on 18 cancer registries in 14 states that collectively represent 26% of the US population, including 23% of African Americans, 40% of Hispanics, 42% of American Indians and Alaska Natives, 53% of Asians, and 70% of Hawaiian/Pacific Islanders.[28] 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,[29] ovarian,[30] and endometrial[31] cancer, and patterns of care studies of the dissemination of advanced cancer treatment modalities throughout different population groups and into community practice.[32–34] With approval, researchers may be granted access to the SEER–Medicare linked data files, which include Medicare claims prior to, during, and following cancer diagnosis and treatment.[35] Topics studied include influences of treatment, facility, and provider characteristics, and interventions on survival and cost outcomes,[36–38] as well as disparities in care.[39]

Rare diseases

Genetic disease registries and other registries that study special conditions provide the ability to understand the long-term natural history of these diseases as well as the effectiveness of various interventions and practices used in their treatment. Registries are also being used for emerging highly infectious diseases, such as highly pathogenic avian influenza. The lysosomal storage disorders (LSDs) are a group of genetic conditions characterized by enzyme deficiencies, leaving 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 natural history, treatment patterns, and efficacy of existing treatments for LSDs including Fabry disease, Gaucher disease, hereditary angioedema, mucopolysaccharidosis I (MPS 1) and 2 (MPS 2, Hunter's syndrome), and Pompe disease have been developed by manufacturers of enzyme replacement therapies and other treatments such as Genzyme and Shire, and by patient associations.[40] 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 analysis of this kind of registry data depends on the degree to which inclusion of subjects in the registry is not biased with regard to treatments and outcomes, and that longitudinal follow up is as complete as possible.

Registries for effectiveness

Treatment effectiveness may be evaluated in registry-based studies, either for single or multiple treatments, with the latter being subject to the same methodologic issues and limitations of other observational studies of comparative effectiveness (see also Chapter 32).[41] For example, the avian influenza registry (http://www.avianfluregistry.org) is designed both to study the clinical course of the disease and treatment effectiveness. Avian influenza (H5N1), a strain of flu originating in birds but transmissible to humans in rare instances usually through direct or indirect contact with poultry, may be highly pathogenic in some strains. A disease registry funded by Hoffmann-La Roche, manufacturers of a popular antiviral, has enrolled hundreds of laboratory-confirmed cases contributed from local treating physicians and health authorities in 12 countries, the largest collection of individual case data ever assembled. Registry data were used to document an approximately 50% reduction in mortality among antiviral treated patients compared with patients who did not receive antiviral treatment.[12]

Device registry studies also commonly evaluate effectiveness or comparative effectiveness as an endpoint. 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.[42]

Safety/regulatory registries

Registries have long had a role in providing information on safety and product usage patterns for marketed products. For example, it is often of interest to regulatory agencies to understand if products are being prescribed appropriately, so registries are sometimes used to collect information about the characteristics of users. Also, since registries may have long periods of follow up, they can be a useful means to study delayed benefits and risks. For example, there are product registries that follow children who used growth hormones to monitor long-term safety (e.g., the Registry of Growth Hormone patients using Norditropin, NCT00615953, sponsored by NovoNordisk). Registries are also an approach that may be used to follow patients who received gene-based treatments in clinical trials as a means of developing information on long-term safety. The US Food and Drug Administration (FDA) has published some general guidance on this topic, which must be tailored to specific features of the treated population and expected risks based on preclinical studies.[43] The European Medicines Agency's Committee for Medicinal Products for Human Use (CHMP) has provided more detail on factors that should be addressed in risk management plans for the category called Advanced Therapy Medicinal Products, which include gene therapy.[44]

In the post-Food and Drug Administration Amendments Act (FDAAA) of 2007, the FDA looks to registries as one approach for sponsors to address certain postmarketing safety commitments. These include product registries, pregnancy registries, and other uses, sometimes as part of REMS programs (see below). In fact, the role of pregnancy registries and "restricted use" (or controlled distribution) product registries in monitoring product safety may be among the most widely known uses of registries (see Chapter 29). 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*.[45] 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 thrombopoeitin 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. A class-wide registry to explore the risk of cancer in children with rheumatic disease treated with tumor necrosis factor inhibitors was also discussed.

Registries are often used in REMS to support safety. REMS are required risk management plans that use risk minimization strategies to go beyond product labeling (see Chapter 29). The tools used in REMS include medical guides, communication plans for health care providers, elements to assure safe use (ETASU), and implementation systems. Registries are used in various aspects of REMS including safety evaluations for high-risk populations such as pregnancy registries for patients using certain treatments of interest. Product registries are also frequently used to characterize product users to assess whether prescribing is occurring consistent with labeling and contraindications, and to assure that users have received appropriate education and testing. Registries may also form the backbone of controlled distribution systems (also referred to as "performance-linked access systems"). For example, the controlled distribution program for clozapine requires regular blood tests to monitor white blood cell counts; the specialty pharmacies cannot distribute clozapine without a valid prescription and updated blood test results. The isotretinoin program requires that recent pregnancy tests are registered as a condition of receiving prescription refills. A restricted distribution program was also recently initiated for romiplostim, a thrombopoeitin mimetic class antihemorrhagic for treatment of thrombocytopenia in patients with ITP manufactured by Amgen.[45] The operational challenges of determining the sampling frame of patients and physicians are often particularly challenging for REMS. Plans often require setting a target for the percentage of the total at-risk population that would need to be included to be sufficient to safeguard against "small" risks. These narrow-purpose registries are rarely used for research purposes, largely due to the constraints under which the data were collected, for example single-purpose informed consents.

The postmarketing surveillance of medical devices may incorporate registry approaches to an even greater extent than the postmarket surveillance of drugs and biologics. Because regulatory approval of medical devices may not always require randomized controlled trials, depending on the nature of the condition to be treated and the device, even more questions about effectiveness and safety may remain after approval (see Chapter 27). Both the effectiveness and safety of devices are dependent, in part, on the skill and experience of the "operator" who inserts, implants, or otherwise administers the device, and often on features of the hospital or interventional setting.[46] Imaging results or other detailed assessments not typically found in administrative data may also be required to assess safety or efficacy outcomes. The level of required clinical detail to adequately monitor device safety and effectiveness may only be obtainable through prospective data collection such as that employed in registries.

Registries of product use in special populations

Registries may focus on special population subgroups of patients who are exposed to a drug or other medical product, such as pregnant women or children, when there is reason for concern regarding the effect of perinatal exposure on the developing embryo or fetus, or when long-term health outcomes in children as a results of product exposure are unknown and need to be monitored.

Pregnancy

Pregnancy registries may include all women who become pregnant (or give birth) within a defined population, or focus on women who become pregnant and are exposed to a medical product (also see Chapter 28.) Like any pharmacoepidemiologic study, studies using product- or drug-class specific pregnancy registries are most useful when they strive to include all exposed women, rather than rely on passive reporting of exposed pregnancies.[47] In fact, registries that are derived from haphazard reporting, rather than systematic reporting, are of low information value for pharmacoepidemiology. Whether the commitment required for registry participation and collection of outcome data may in any way bias the enrollment toward a select group

of health-conscious women should also be considered in registry planning and design.

Both the US FDA and EMA CHMP have published guidances regarding the design of pregnancy registries to provide postmarketing safety data. The FDA guidance defines pregnancy exposure registries as follows: "A pregnancy exposure registry is a prospective observational study that actively collects information on medical product exposure during pregnancy and associated pregnancy outcomes."[47] It goes on to emphasize as strengths of this approach the prospective nature of a pregnancy exposure registry, with collection of exposure data before pregnancy outcomes are known, and the corresponding ability to collect information on multiple outcomes, and draws a distinction between pregnancy exposure registries and pregnancy prevention programs.

The CHMP guidance describes several specific roles that registries may play in providing important postauthorization safety data regarding perinatal exposures. These include: population-based birth defect registries, which may be used to conduct case–control studies of a range of perinatal exposures, prospective pregnancy registries, and linkage of registry data with other data sources to provide information on long-term "structural" defects or cancers that may occur years after the perinatal exposure.[48] The Swedish National Birth Register is one of the largest pregnancy exposure registries, collecting data on all drug exposures in pregnant women since 1973 with 99% registration,[49] and has served as the population base for numerous studies of outcomes associated with perinatal exposures. Published pharmacoepidemiologic analyses of the Swedish National Birth register in recent years include studies of antiepileptic drug exposure and head circumference,[50] and studies of maternal use of many drugs and drug classes such as thyroid hormones,[51] antipsychotics and lithium,[52] loperamide,[53] and antiasthmatic drugs.[54] It has been linked to the Swedish Cancer Registry to study birth outcomes after *in vitro* fertilization,[55] and maternal exposures including oral contraceptive use and childhood brain tumors.[56]

The Antiretroviral Pregnancy Registry is an example of a pregnancy registry designed to assess risks associated with all antiretroviral medications taken by women during pregnancy.[57] Started in 1989, the registry has studied pregnancy outcomes associated with a long and changing array of treatments as newer therapies and multidrug regimens have become available. Strengths of the registry include the open inclusion criteria, large number of enrolled pregnancies, and global reach. A limitation of the registry is its reliance on passive reporting of pregnancies, so that it is difficult to determine how representative the data are of all women treated with antiretrovirals in pregnancy.[47] As of recent reports,[58,59] no increased risk of birth defects has been associated with antiretroviral treatment overall or for the specific medications abacavir, atazanavir, efavirenz, emtricitabine, indinavir, lopinavir, nelfinavir, nevirapine, ritonavir, stavudine, and tenofovir, for which sufficient numbers of live births (>200) following first-trimester exposures have been monitored to allow detection of a twofold increase in risk of birth defects.

iPledge is a program that combines a pregnancy prevention program with a pregnancy exposure registry for women who become pregnant during or within 30 days of past use of isotretinoin, a known teratogen used in the treatment of acne. The drug is approved for restricted distribution under the condition that women of childbearing potential testify to use of two methods of birth control and undergo monthly pregnancy testing recorded in a database linked to pharmacies, which require approval to dispense the drug. The US manufacturers of isotretinoin jointly provide funding and oversight of the iPledge program. This mandatory registry is a strategy for REMS (see above).

Children

Concern regarding long-term health outcomes such as development of cancer among children who are exposed to medical products, based on theoretical risks, preapproval data, or spontaneous reports, may prompt product manufacturers to conduct registry studies of treated children, often in response to regulatory requirements. Such studies may be designed to follow the children prospectively for many years through a sufficient time period to evaluate the risks associated with product

exposure, or to collect exposure and other covariates of interest with the intent of linking to a cancer registry or other data source at a later time.

For example, a prospective 10-year observational registry of children who use pimecrolimus for treatment of eczema has been established by the manufacturer, Novartis, in response to regulatory commitment to the FDA and is ongoing (ClinicalTrials.gov identifier NCT00568997). The primary outcome is lymphoma, and rates of lymphoma and any systemic malignancy among registry subjects are to be compared to those estimated by the SEER registry of the US National Cancer Institute (NCI).

A few pediatric registries of individual anti-TNF products in multiple therapeutic areas are underway or have been completed, such as infliximab for inflammatory bowel disease (NCT00606346) and etanercept for psoriasis (NCT01100034), while a consolidated pediatric rheumatologic disease registry focused on the safety and effectiveness of newer as well as traditional treatments for juvenile arthritis and other conditions is still under discussion at FDA.[60] Also, postmarket commitments to continue to follow participants in previous clinical trials by manufacturers of these products have also been made.[61] The limitations of studying individual products include lack of information on patients who switch therapies or use multiple therapies, smaller study sizes, and differences in inclusion and exclusion criteria, which may make comparisons of products across studies difficult.

Registries for quality improvement

Registries can provide information that helps to shape clinical trials and to understand medical practice. For example, a recent registry of physician certification and outcomes for patients receiving implantable cardioverter-defibrillators showed that non-electrophysiologists implant nearly one-third of ICDs in the US. A significant finding of the study was that implantations by a non-electrophysiologist were associated with a higher risk for procedure complications and a lower likelihood of receiving a CRT-D device when indicated, compared with patients whose ICD was implanted by an electrophysiologist.[62]

The increasing use of web-based data collection systems for registries provides access to reporting and analytic tools to participating physicians and patients, or to the broader public. The ability to provide direct value to participants in a registry, through quality and performance metrics, individual patient treatment and clinical summaries, or surveillance statistics, may play an important role in maintaining the engagement of sites and patients with the registry and with increased adherence to accepted quality guidelines. The *Get With the Guidelines* program for coronary artery disease is an example of a web-based, point-of-care registry that was used to reinforce evidence-based guidelines for coronary artery disease. The effects of using evidence-based guidelines and reminders were evaluated in more than 100,000 patients from 500 hospitals. The evidence accumulated by the registry showed that the program resulted in increased use of statins (by more than 80% in patients where statin use was indicated), and a doubling of referrals to smoking cessation programs.[63]

Registries for reimbursement/coverage decisions

Registries are also used to study medical decision making. For example, the US Centers for Medicare and Medicaid Services (CMS) was interested in extending coverage for the use of positron emission tomography (PET) scanning in diagnosing certain cancers. They used a registry approach during which they agreed to provide reimbursement for PET scanning in these situations of interest in exchange for information about how the physicians used this information in medical care. The results of the registry showed that the information was indeed useful and meaningfully reduced the need for biopsies and influenced the course of care. As a result of this registry, additional reimbursement coverage was granted for these previously not covered uses.[64]

The Future

The current use of registries as research data repositories reflects the increasingly visible roles they

play in understanding the natural history and treatment of disease, studying a vast range of treatment outcomes, and meeting postmarketing commitments (including REMS programs) for medical product manufacturers. The future looks to bring further emphasis on registries and linked registries as auxiliary or primary components of national and international product safety monitoring systems, and for ongoing surveillance of care quality and patient outcomes.[65,66] Studies derived from registries are beginning to incorporate the same advanced methods as other pharmacoepidemiologic studies in their design and analysis, such as new user designs, and methods to address missing data, such as multiple imputation.

Registries will continue to be used for safety, and may play a meaningful role in the ongoing development of FDA's Sentinel Initiative (see Chapter 30), which foresees the inclusion of registries as a secondary data source within this system for proactive safety monitoring of regulated products:[67]

> A national electronic system that will transform FDA's ability to track the safety of drugs, biologics, medical devices—and ultimately all FDA-regulated products once they reach the market—is now on the horizon. Launched in May 2008 by FDA, the Sentinel Initiative aims to develop and implement a proactive system that will complement existing systems that the Agency has in place to track reports of adverse events linked to the use of its regulated products.
>
> Monitoring the safety of its regulated products is a major part of FDA's mission to protect public health. The Sentinel System would enable FDA to actively query diverse automated health care data holders—like electronic health record systems, administrative and insurance claims databases, and registries—to evaluate possible medical product safety issues quickly and securely.
>
> Sentinel will be developed and implemented in stages. As the system is envisioned, data would continue to be managed by its owners and questions would be sent to the participating data holders. Within pre-established privacy and security safeguards, these data holders would evaluate their information and send summary results to FDA. It is also anticipated that Sentinel will facilitate the development of active surveillance methodologies related to signal detection, strengthening, and validation.

Though the distributed data network currently envisioned for the Sentinel System would draw mainly from electronic health record and administrative claims data for signal detection, it is anticipated that links to external registry data will be required at times to be able to study questions of device safety, as well as to incorporate immunization records or birth and death records to provide detail of exposures that are not always available in electronic health records or claims, or to link exposures or outcomes that may occur outside the limited periods of enrollment within a given health plan in the current system.[68]

Registries are going to be particularly important for studying comparative effectiveness (see Chapter 32) and it is likely that most safety studies will also incorporate internal comparators to facilitate studies of comparative effectiveness. The Institute of Medicine's 100 initial priority topics for comparative effectiveness research include many topics that may be studied using registries.[69] For example, a registry of people being treated for hearing impairment might be an ideal approach for: "comparing the effectiveness of different treatments (e.g., assistive listening devices, cochlear implants, electric-acoustic devices, habilitation and rehabilitation methods [auditory/oral, sign language, and the total communication]) for hearing loss in children and adults, especially individuals with diverse cultural, language, medical and developmental backgrounds."

We expect to see further evolution of systems for assembling and linking different sources of information for registries, which will facilitate pharmacoepidemiologic research and pharmacovigilance. Internet-based data collection methods are increasingly used to facilitate multicenter and international studies, especially of rare diseases (e.g., lysosomal storage disorders such as Pompe disease[70]). While truly novel data sources and methods of data acquisition may arise, efforts are underway on various fronts for new uses and expanded uses of existing data sources, such as integration with electronic medical records. Much remains to be learned from direct contact with patients and care providers. Newer methods of collecting data directly from patients and from doctors, including internet-based data collection and integrated voice response systems with autocoding fea-

tures, will facilitate rapid accumulation of data; however, the validity, accuracy, and usefulness of these new methods need to be evaluated. For example, one work package in the European PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics) Consortium, funded by the Innovative Medicines Initiative, is testing the feasibility, efficiency, and usefulness of broad-scale, direct collection of data from pregnant women on use of prescription and non-prescription medications, and herbals.[71] This study will involve multilingual data collection through the internet and using interactive voice response systems.

Registries are expected to continue adopting newer methods for observational study design and analysis, and incorporating new technologies to accomplish objectives such as facilitated safety reporting. For example, concepts of multilevel analysis and study design[72–74] are applicable to multisite or multiprovider studies where it is expected that clinic/hospital or provider/operator-level differences, in addition to product exposure, may influence patient outcomes.[46,75] Specific safety or comparative effectiveness hypotheses may be combined with broader descriptive aims by "nesting" these analyses within larger product registries, drawing on propensity score methods for achieving balanced comparisons.

Enhancements to technology for existing electronic systems will facilitate more rapid and complete adverse drug event reporting, which is often a built-in feature of safety registries. ASTER (Adverse Drug Event [ADE] Spontaneous Triggered Electronic Reports) is a project at Partners Healthcare (within the local electronic health-record system) that brings up MedWatch forms (see Chapter 10) on screen to be completed and submitted to the FDA in response to specific event "triggers."[76] The impact in limited studies so far is a reduction in the burden of reporting on the physician and an increase in the reporting rate. The broad application to electronic medical record systems to registries will depend on adoption of interoperability standards and use of common data standards.

The use of registries will likely continue to grow in coming years to support the increasing demand for real-world data to meet postmarketing study commitments and to inform treatment and reimbursement decisions.

References

1. Graunt J. *Natural and Political Observations Mentioned in a Following Index and Made Upon the Bills of Mortality*. New York: Arno Press, 1975. [Reprint of the 1662 edition.]
2. Vandenbroucke J. Observational research, randomised trials, and two views of medical science. *PLOS Med* 2008; **5**: 339–43.
3. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies and the hierarchy of research designs. *N Engl J Med* 2000; **342**: 1887–92.
4. Rawlins, MD, De Testimonio. On the evidence for decisions about the use of therapeutic interventions. *Clin Med* 2008; **8**: 579–88.
5. Avorn J. In defense of pharmacoepidemiology—embracing the yin and yang of drug research. *N Engl J Med* 2007; **357**: 2219–21.
6. Gliklich RE, Dreyer NA, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*, 2nd edn. (Prepared by Outcome DEcIDE Center [Outcome Sciences, Inc. dba Outcome] under Contract No. HHSA29020050035I TO1.) AHRQ Publication No. 10-EHC049. Rockville, MD: Agency for Healthcare Research and Quality, 2010.
7. Mahon BE, Shea KM, Dougherty NN, Loughlin AM. Implications for registry-based vaccine effectiveness studies from an evaluation of an immunization registry: A cross-sectional study. *BMC Public Health* 2008; **8**: 160.
8. Samuels RC, Appel L, Reddy SI, Tilson RS. Improving accuracy in a computerized immunization registry. *Ambul Pediatr* 2002; **2**: 187–92.
9. Wood D, Saarlas KN, Inkelas M, Matyas BT. Immunization registries in the United States: implications for the practice of public health in a changing health care system. *Annu Rev Public Health* 1999; **20**: 231–55.
10. Zaroff JG, diTommaso DG, Barron HV. A risk model derived from the National Registry of Myocardial Infarction 2 database for predicting mortality after coronary artery bypass grafting during acute myocardial infarction. *Am J Cardiol* 2002; **90**: 1–4.
11. Cystic Fibrosis Foundation. *Patient Registry Report*. www.cff.org/LivingWithCF/QualityImprovement/PatientRegistryReport (accessed May, 2010).

12. Adisasmito W, Chan PKS, Lee N, Oner AF, Gasimov V, Aghayev F, *et al.* Effectiveness of antiviral treatment in human influenza H5N1 infections: analysis from a global patient registry. *J Infect Dis* 2010; **202**: 1154–60.
13. Oner AF, Bamgboye E, Gasimov V, Zaman M, Chan PKS, Lee N, *et al. Avian Influenza Infections in Children: Symptoms and Treatment*. European Society for Paediatric Infectious Diseases, Nice, France, May, 2010.
14. 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.
15. Herbst AL, Kurman RJ, Scully RE. Clear-cell adenocarcinomas of the genital tract in young females. Registry report. *N Engl J Med* 1972; **287**: 1259–64.
16. Charrow J, Andersson HC, Kaplan P, Kolodny EH, Mistry P, Pastores G, *et al.* The Gaucher Registry: demographics and disease characteristics of 1698 patients with Gaucher disease. *Arch Intern Med* 2000; **160**: 2835–43.
17. Friedberg JW, Taylor MD, Cerhan JR, Flowers CR, Dillon H, Farber CM, *et al.* Follicular lymphoma in the United States: First report of the national lymphocare study. *J Clin Oncol* 2009; **27**: 1202–8.
18. Massop D, Dave R, Metzger C, Bachinsky W, Solis M, Shah R, *et al.* Stenting and angioplasty with protecting in patients at high-risk for endarterectomy: SAPPHIRE worldwide registry first 2,001 patients. *Catheter Cardiovasc Interv* 2009; **73**: 129–36.
19. Xian Y, Pan W, Peterson ED, Heidenreich PA, Cannon CP, Hernandez AF, *et al.* Are quality improvements associated with the Get With the Guidelines-Coronary Artery Disease (GWTG-CAD) program sustained over time? A longitudinal comparison of GWTG-CAD hospitals versus non-GWTG-CAD hospitals. *Am Heart J* 2010; **159**: 207–14.
20. Illinois Department of Public Health. *Illinois Comprehensive Automated Immunization Registry Exchange (I-CARE) Fact Sheet*. http://www.idph.state.il.us/health/vaccine/icarefs.html (accessed Aug., 2010).
21. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**: 200–8.
22. Rothman KH, Greenland S, Lash TL. *Modern Epidemiology*, 3rd edn. New York: Lippincott Williams and Wilkins, 2008.
23. Von Elm E, Altman DG, Egger M, Pocock SJ, Gotzche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; **147**: 573–7.
24. Dreyer NA, Schneeweiss S, McNeil B, Berger ML, Walker A, *et al*, on behalf of the GRACE Initiative. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care* 2010; **16**: 467–71.
25. Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES. Variation in carotid endarterectomy mortality in the medicare population: trial hospitals, volume, and patient characteristics. *JAMA* 1998; **279**: 1278–81.
26. Kaufman PA, Dreyer NA, Mayer M, Sloan JA, Chung CY, Sing A, *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*. American Society of Clinical Oncology, Chicago, June, 2010.
27. National Institute of Arthritis and Musculoskeletal and Skin Diseases, US Department of Health and Human Services. *Patient Research Registries*. http://www.niams.nih.gov/Funding/Funded_Research/registries.asp (accessed May, 2010).
28. National Cancer Institute. *Surveillance, Epidemiology, and End Results. Overview of the SEER Program*. http://seer.cancer.gov/about/ (accessed May, 2010).
29. 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.
30. Anon. Oral contraceptive use and the risk of ovarian cancer. *JAMA* 1983; **249**: 1596–9.
31. Anon. Oral contraceptive use and the risk of endometrial cancer. *JAMA* 1983; **249**: 1600–4.
32. Shavers VL, Harlan LC, Jackson M, Robinson J. Racial/ethnic patterns of care for pancreatic cancer. *J Palliat Med* 2009; **12**: 623–30.
33. 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**: 577–83.
34. 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.
35. National Cancer Institute. *Surveillance, Epidemiology, and End Results Program. SEER-Medicare Linked*

Database. http://seer.cancer.gov/about/seermedicare.html (accessed May, 2010).
36. 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**: 1054–63; discussion 1064.
37. 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**: 3848–57.
38. Breslin TM, Morris AM, Gu N, Wong SL, Finlayson EV, Banerjee M, *et al.* Hospital factors and racial disparities in mortality after surgery for breast and colon cancer. *J Clin Oncol* 2009; **27**: 3945–50.
39. Rouhani P, Arheart KL, Kirsner RS. Differences in melanoma outcomes among Hispanic Medicare enrollees. *J Am Acad Dermatol* 2010; **62**: 768–76.
40. US Hereditary Angioedema Association. *US HAE Scientific Registry: Patient Driven Research for a Cure*. http://www.haea.org/get-involved/us-haea-scientific-registry/ (accessed June 2011).
41. Dreyer NA, Garner S. Registries for robust evidence. *JAMA* 2009; **302**: 790–1.
42. Douglas PS, Brennan JM, Anstrom KJ, Sedrakyan A, Eisenstein EL, *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**: 1629–41.
43. Center for Biologics Evaluation and Research (CBER), *Food and Drug Administration, United States Department of Health and Human Services. Guidance for Industry: Gene Therapy Clinical Trials—Observing Subjects for Delayed Adverse Events*. http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/CellularandGeneTherapy/ucm072957.htm (accessed May, 2010).
44. Committee for Medicinal Products for Human Use (CHMP). *Guideline on Safety and Efficacy Follow-Up—Risk Management of Advanced Therapy Medicinal Products*, Doc. Ref. EMEA/149995/2008 . European Medicines Agency, 2008.
45. Hobbs MN. Registries rising: FDA looking at TNF inhibitors; AHRQ updates standards. *Pink Sheet* 2009; **71**: 8–9.
46. Sedrakyan A, Marinac-Dabic D, Normand SL, Mushlin A, Gross T. Framework for evidence evaluation and methodological issues in implantable device studies. *Med Care* 2010; **48** (6 Suppl): S121–8.
47. Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER). *Guidance for Industry: Establishing Pregnancy Exposure Registries*. Food and Drug Administration, United States Department of Health and Human Services. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071639.pdf (accessed April, 2010).
48. Committee for Medicinal Products for Human Use (CHMP). *Guideline on the Exposure to Medicinal Products During Pregnancy: Need for Post-Authorisation Data*, EMEA/CHMP/313666/2005. European Medicines Agency, Adopted November 2005. http://www.ema.europa.eu/pdfs/human/phvwp/31366605en.pdf (accessed April, 2010).
49. Cnattingius S, Eriksson A, Gunnarskog J, Källén B. A quality study of a medical birth registry. *Scand J Soc Med* 1990; **18**: 143–8.
50. Almgren M, Källén B, Lavebratt C. Population-based study of antiepileptic drug exposure in utero—influence on head circumference in newborns. *Seizure* 2009; **18**: 672–5.
51. Wikner BN, Sparre LS, Stiller CO, Källén B, Asker C. Maternal use of thyroid hormones in pregnancy and neonatal outcome. *Acta Obstet Gynecol Scand* 2008; **87**: 617–27.
52. Reis M, Källén B. Maternal use of antipsychotics in early pregnancy and delivery outcome. *J Clin Psychopharmacol* 2008; **28**: 279–88.
53. Källén B, Nilsson E, Otterblad Olausson P. Maternal use of loperamide in early pregnancy and delivery outcome. *Acta Paediatr* 2008; **97**: 541–5.
54. Källén B, Otterblad Olausson P. Use of anti-asthmatic drugs during pregnancy. 1. Maternal characteristics, pregnancy and delivery complications. *Eur J Clin Pharmacol* 2007; **63**: 363–73.
55. Nygren KG, Finnström O, Källén B, Olausson PO. Population-based Swedish studies of outcomes after in vitro fertilisation. *Acta Obstet Gynecol Scand* 2007; **86**: 774–82.
56. Linet MS, Gridley G, Cnattingius S, Nicholson HS, Martinsson U, Glimelius B, *et al.* Maternal and perinatal risk factors for childhood brain tumors (Sweden). *Cancer Causes Control* 1996; **7**: 437–48.
57. Covington DL, Tilson H, Elder J, Doi P. Assessing teratogenicity of antiretroviral drugs: monitoring and analysis plan of the Antiretroviral Pregnancy Registry. *Pharmacoepidemiol Drug Saf* 2004; **13**: 537–45.
58. Antiretroviral Pregnancy Registry Steering Committee. *Antiretroviral Pregnancy Registry International Interim Report for 1 January 1989 through 31 July 2009*. Wilmington, NC: Registry Coordinating Center, 2009. http://www.APRegistry.com (accessed June 2011).

59. Watts DH, Covington DL, Beckerman K, Garcia P, Scheuerle A, Dominguez K, *et al.* Assessing the risk of birth defects associated with antiretroviral exposure during pregnancy. *Am J Obstet Gynecol* 2004; **191**: 985–92.
60. Center for Drug Evaluation and Research (CDER). *Public Workshop on Developing a Consolidated Pediatric Rheumatology Observational Registry, May 12–13, 2009*. http://www.fda.gov/downloads/Drugs/NewsEvents/UCM171762.pdf (accessed May, 2010).
61. Food and Drug Administration, United States Department of Health and Human Services. *Postmarket Requirements and Commitments*. http://www.accessdata.fda.gov/scripts/cder/pmc/index.cfm (accessed May, 2010).
62. Curtis JP, Luebbart JJ, Wang Y, Rathmore SS, Chen J, Heidenreich PA, *et al.* Association of physician certification and outcomes among patients receiving an implantable cardioverter-defibrillator. *JAMA* 2009; **301**: 1661–70.
63. LaBresh KA, Gliklich R, Liljestrand J, Peto R, Ellrodt AG. Using "get with the guidelines" to improve cardiovascular secondary prevention. *Jt Comm J Qual Saf* 2003; **29**: 539–50.
64. Hillner BE, Siegel BA, Liu D, Shields AF, Gareen IF, Hanna L, *et al.* Impact of positron emission tomography/computed tomography and positron emission tomography (PET) alone on expected management of patients with cancer: initial results from the National Oncologic PET registry. *J Clin Onc* 2008; **26**: 5316–19.
65. Krishnan JA, Mularski RA. Acting on comparative effectiveness research in COPD. *JAMA* 2010; **303**: 2409–10.
66. Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox PW, Gliklich RE. Why observational studies should be among the tools used in comparative effectiveness research. *Health Affairs* 2010; **29**: 1818–25.
67. Food and Drug Administration, United States Department of Health and Human Services. *FDA's Sentinel Initiative: Transforming How we Monitor the Safety of FDA-related Products*. http://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm (accessed May, 2010).
68. 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.
69. Institute of Medicine. *100 Initial Priority Topics for Comparative Effectiveness Research*. http://www.iom.edu/~/media/Files/Report%20Files/2009/ComparativeEffectivenessResearchPriorities/Stand%20Alone%20List%20of%20100%20CER%20Priorities%20-%20for%20web.ashx (accessed May, 2010).
70. Genzyme. Pompe Registry. https://www.lsdregistry.net/pomperegistry (accessed June, 2010).
71. European Medicines Agency. PROTECT. http://www.imi-protect.eu (accessed May, 2010).
72. Merlo J, Chaix B, Yang M, Lynch J, Rastam L. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Comm Health* 2005; **59**: 443–9.
73. Diez-Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* 2000; **21**: 171–92.
74. Leyland AH, Goldstein H, eds. *Multilevel Modeling of Health Statistics*. Wiley, 2001.
75. Gort M, Otter R, Plukker JTM, Broekhuis M, Klazinga NS. Actionable indicators for short and long term outcomes in rectal cancer. *Eur J Cancer* 2010; **46**: 1808–14.
76. Arnold M. Pfizer seeks ways to ease adverse events reporting digitally. *Medical Marketing and Media*, March 30, 2010 http://www.mmm-online.com/pfizer-seeks-ways-to-ease-adverse-events-reporting-digitally/article/166947/ (accessed May, 2010).

CHAPTER 22
Field Studies

David W. Kaufman
Slone Epidemiology Center at Boston University, Boston, MA, USA

Introduction

Epidemiologic studies in which data are collected in the field for the purpose of evaluating a specific hypothesis are known as "field" or "*ad hoc*" studies. These are to be contrasted with studies that use pre-existing data, principally from health-care databases. Much of pharmacoepidemiology today consists of the latter "database" studies (see Part III Section B). While such studies have advantages of time efficiency, cost, and some validity benefits, there are potential drawbacks in terms of subject definitions and data availability, and there may not be appropriate data available to address particular questions. With the ability to tailor subject enrollment and data collection to a specific research question, field studies continue to play an important role in pharmacoepidemiology. It is also worth noting that until the relatively recent advent of databases, virtually all epidemiology was conducted by means of field studies.

All types of epidemiologic investigations, including cohort, case–control, and cross-sectional designs, can be conducted as field studies, as long as subject enrollment and data collection are part of the process. Postmarketing randomized trials can also be considered to be field studies—these are described in Chapter 36 and will not be covered here. A special type of field study is case–control surveillance, which is described in Chapter 19 and also will not be covered here.

Field studies are by their nature more expensive and much slower than analyses of existing data, which raises the question, why conduct them at all in the modern world of tight deadlines and databases that are readily available? The general answer is that there are situations where a field study is the only way to recruit the subjects and/or obtain the information needed to provide a valid answer to a specific research question. This chapter is devoted to explaining the details of that deceptively simple statement.

Strengths

The strengths and weaknesses of field studies are generally opposite to those of database studies. On the strength side, it is often possible to more rigorously define outcomes, it may be more feasible to enroll subjects with very rare conditions, and it is especially more feasible to obtain the information needed to study questions for which administrative data are inadequate, because the study infrastructure and data collection procedures are set up specifically for the purpose of accomplishing these goals.

Outcome definition

A general problem with databases is that outcomes are defined by diagnosis codes, sometimes augmented by more detailed clinical information, for

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

example with HMO databases such as Kaiser Permanente.[1] Such codes are frequently insufficient for confirming the validity of diagnoses. An example is Stevens–Johnson syndrome, where a detailed review of clinical information, which ideally extends to evaluating photographs of patients' lesions, is necessary to ensure a valid series of cases for epidemiologic study.[2] In database studies, the choice is to create an algorithm based on diagnosis codes and perhaps some treatment information, or better, to obtain access to patient records for the needed information, which can be a difficult process unless electronic medical records are available.[3] Even if an algorithm is used, it is generally advisable to conduct a validation study based on medical records in a sample of patients. In contrast, collection of the needed information can be built into the protocol of a field study. Then the appropriate ongoing review process can be established (see the Particular Applications section for further details).

Studying extremely rare diseases

While databases cover large numbers of subjects, some diseases are so rare that until recently even the largest databases were insufficient to produce enough cases for informative study. In these situations, it has been necessary to set up large population-based case finding networks to enroll the patients. Examples are agranulocytosis, aplastic anemia, and Stevens–Johnson syndrome/ toxic epidermal necrolysis (SJS/TEN), conditions that are of frequent interest because they are often induced by drugs and disproportionately result in regulatory or legal action.[4,5] The International Agranulocytosis and Aplastic Anemia Study (IAAAS),[6,7] conducted in Israel and Europe in the 1980s, covered a population of 23 million subjects over a period of 6 years to prospectively enroll 270 cases of agranulocytosis and 154 of aplastic anemia. The data collection network of the Severe Cutaneous Adverse Reaction (SCAR) study[8,9] covered a population of approximately 130 million in four European countries for 6 years to enroll 373 cases of SJS/TEN; a continuation of that project, the EuroSCAR study,[10] covered a similar population in five countries. It would have been difficult to provide such a large population experience in any other way, although with the use of large government databases, especially the availability of Medicare data (see Chapter 14), and the linking of multiple databases (see Chapter 12), that is becoming feasible; the future will see linked databases exceeding 100 million individuals (see Chapter 30).

Exposure and covariate information tailored to the research question

Setting up a data collection system specifically to study a particular question has the major advantage of allowing the collection of precisely the information needed, which frequently can only be obtained from the subjects themselves. While databases have the advantage of prescription records that are independent of any research agenda, drugs are not always taken as prescribed, and prescription drugs can be obtained from other sources, such as friends and relatives. If non-prescription (OTC) drug use is relevant, that information can usually only be obtained directly from the subjects. The same is true of herbals and other supplements. Details of habits such as alcohol and tobacco consumption, and patient-reported outcomes such as quality of life (see also Chapter 39), generally require access to the subjects themselves. Other sources of information, such as medical records and medical care providers, can also be accessed in a field study to provide important details that may not otherwise be available, such as specifics of oncology treatment regimens beyond the simple fact of prescriptions. The value of obtaining detailed information from records and providers for the rigorous definition of diseases under study has already been noted. The collection of appropriate and specifically relevant exposure and covariate information is also one of the signal advantages of field studies.

Weaknesses

Time, cost, and logistics

Because of the need to set up and maintain data collection networks, enroll subjects, and obtain data, field studies are time consuming and expensive. A prospective follow-up study, in particular,

usually requires actually following the individual subjects for a period of years.[11–14] While *ad hoc* case–control studies are generally faster than cohort studies, with exposures of interest occurring prior to enrollment, the prospective enrollment of incident cases as they occur can also take years,[7,9,15,16] depending on the incidence of the disease under study. Many questions in pharmacoepidemiology are urgent in nature, especially if driven by regulatory concerns; thus, the long lead-time required to conduct a field study can be a real barrier that must be balanced against the requirement for information that cannot be obtained by other approaches.

The corollary to the time requirements of field studies is that a substantial staff is usually needed to manage the logistics of a large-scale enrollment network and the subject by subject data collection, which are not issues with database studies. Some of the specific logistical challenges may include the following:

- identifying and contacting the study subjects;
- overcoming language barriers and other communication problems in international and other studies with many centers and co-investigators;
- hiring, training, and monitoring the field staff;
- tracking a potentially large flow of study-related materials, including Institutional Review Board (IRB) approvals from each site, consent forms, completed questionnaires, and abstracts of medical records. There are also issues of storage of this volume of materials and secure storage of sensitive patient information.

Personnel costs are always by far the largest component of a field study budget, which can frequently run to millions of dollars. This can be a practical barrier to the conduct of these studies, particularly with government sources of funding. For industry sponsors, it is more a matter of how important they (or drug regulatory agencies) deem the question. The temptation to pursue faster but less valid results is ever present, although that can lead to inadequate answers to important questions.

Issues of study validity

Avoidance of selection bias

Although more precise inclusion criteria are usually possible compared to database studies, it is still necessary to identify and enroll most subjects who meet the criteria to avoid selection bias, and this can present a substantial challenge. An effective general strategy to maximizing subject identification is to avoid more passive approaches such as voluntary referrals from clinicians or self-selection by subjects, and instead employ "active ascertainment."[7,17] Otherwise, one risks referral bias. Further, when eligible subjects are identified, assiduous efforts must be made to recruit them to ensure a reasonable participation rate. These are substantial challenges, requiring well-trained and dedicated study staff, and, in recent years, participation in even well-conducted epidemiologic studies has declined.[18] Recruitment rates and the potential for non-response bias must always be considered in judging the validity of field studies.

Avoidance of information bias

The other major validity concern that is more particularly a problem for field studies, as opposed to analyses of pre-existing data collected for purposes other than research, is information bias. This can apply either to the identification of outcomes among exposed and non-exposed cohorts, or the collection of exposure information in case–control studies. It can arise because the process is not consistent for all subjects, because of conscious or unconscious bias on the part of study staff, or different data collection procedures that result from systemic differences in the medical care system for individuals with different illnesses. Another source of information bias is differential recall of information by study subjects according to their exposure or disease status (i.e., recall bias). In *ad hoc* case–control studies, where information about drug exposures and potential confounding factors is often obtained by interview, the general concern is that cases may tend to remember and report their histories more completely than controls, who are often relatively healthy and lacking a reason to search their memory for events that explain an illness.[7] A variant of the problem is that individuals who engage in unhealthy behavior, such as excessive use of drugs that may be harmful, may tend to underreport it.[19]

However it arises, information bias can lead either to over- or under-estimation of epidemiologic measures of association. While information bias can never be ruled out, it can be minimized by good practices, including rigorous training of data collection staff, careful design of questionnaires to maximize recall, and procedures that ensure consistent data collection. These important aspects of the conduct of field studies are discussed in more detail later in this chapter.

Other methodological issues

Other potential problems, such as confounding, affect all pharmacoepidemiologic studies and cannot be considered particular weaknesses of field studies. Indeed, such studies often have the ability to obtain more appropriate and detailed information on potential confounding factors. These more generally applicable issues will not be covered in any detail here.

Particular applications

This section describes some of the practical aspects of field studies, taking a "life cycle" approach that covers design, setup, and conduct. Issues related to analysis are generally not unique to field studies, but are touched upon briefly.

Design

What kind of study and data are needed?

The first design consideration is a careful understanding of the study question, which in turn determines the most appropriate approach to providing an informative answer. While this is theoretically always necessary for good study design, it is especially relevant in the case of field studies, where there are fewer conceptual limitations than in studies based on pre-existing data, since each field study starts from a blank sheet. Relevant issues that determine design choices include:

- *Incidence of the outcome of interest*—rare outcomes, such as acute hematological or cutaneous reactions with an annual incidence of a few cases per million,[6,17] generally require a case–control approach to enroll enough cases, whereas more common outcomes, such as myocardial infarction, with an incidence measured in cases per thousand,[20] can be studied by either a case–control or follow-up (cohort) approach (see Chapters 3 and 4 for more detail).
- *Frequency of exposure to the relevant drug*—uncommonly used drugs generally must be studied with a follow-up approach in which the primary criterion for enrollment focuses on users and nonusers. It is obvious that at some point, the combination of a rare outcome with a rare exposure reaches the level where it becomes infeasible for reasons of logistics or cost to conduct an informative field study (see Chapter 4 for more detail).
- *Nature of the putative association*—is it large in relative terms? This influences the sample size requirements (see Chapter 4). Is it an acute effect that occurs soon after exposure, or are latent intervals possible? Does it occur early in treatment with the drug or is a substantial length of exposure required? These latter questions indicate the relevant time window for identifying exposure and outcome, and also determine the level of detail that must be obtained. A study of acute effects may also require considerable detail about the clinical onset of the outcome, in order to properly discern the temporal sequence between exposure and outcome.[6,8,21,22]
- *Where the study should be conducted*—political/ regulatory considerations may dictate the need for data from specific countries.[23] Obviously, the study must be conducted in regions where both the exposure and outcome of interest occur. The exposure side of the equation is particularly relevant in pharmacoepidemiology, since not all drugs are available, or commonly used, in all countries.
- *Sources of information needed to rigorously define the outcome of interest*—is a diagnostic label sufficient? Can patient reports be relied upon? Are medical records or provider reports needed? Is the diagnosis sufficiently inconsistent that a separate review process should be established to ensure uniformity?
- *Sources of information needed to characterize exposure*—patient reports, medical records, healthcare providers? The sources will help to determine the nature of the data collection process, including such issues as self-administered questionnaires

versus personal interviews, having non-study medical staff complete abstraction forms to provide relevant data from records versus interviewing physicians versus requesting the records for abstraction by study staff.

• *Likely confounding that will need to be addressed*—understanding this issue will help determine the information needed on potential confounding factors and other relevant covariates, and the analytic strategy. In turn, the information requirements will determine the sources that need to be tapped to obtain covariate details. Frequently, the best source will be the study subjects themselves.

Protocol development

Once the basic design questions have been answered, the detailed study protocol can be developed. The first key point in this process is to determine what is required to enroll subjects who meet the inclusion criteria, both in terms of the specific infrastructure and the size of the data collection network needed to meet the sample size goals. Then the actual data collection procedures can be specified, a process in which practical considerations play a crucial role, since the protocol has to be not only rigorous, but also feasible. For example, it may be desirable to obtain blood samples for the extraction of DNA among non-hospitalized subjects, but this requires them to come to a site where blood can be drawn, or a home visit by a trained phlebotomist. An alternative that does not yield as much DNA, but is sufficient for many purposes, is the use of buccal swabs or saliva samples, which can be collected by the subjects themselves at home and shipped in prepaid mailers to the study office or laboratory.[24] If data are to be obtained directly from study subjects, a key question is whether this should be done by interview (in-person or via telephone) or self-administered questionnaire (e.g., completing a mailed or online form). In general, interviews allow for more control over the consistency of the data collection process, since it is guided by the interviewer, who can be trained and supervised. Self-administered questionnaires are less expensive since an interviewing staff is not required, but they rely on individual interpretation of the questions by subjects. The anonymity of filling out a questionnaire in privacy can be an advantage for sensitive topics, although it can be argued that the rapport developed by an experienced interviewer is also beneficial in this regard. Practical considerations of cost and feasibility generally dictate the approach chosen.

Data collection instruments

Because a study is only as informative as the data that are recorded, developing instruments that allow for the collection of the needed items is a critical part of the design process. Depending on the details and source of the data to be collected, the instruments may include case report and record abstraction forms, self-administered questionnaires, and interview questionnaires. In all instances, the instruments should be clear and set up for ease of completion.

Case report forms (CRFs) which are filled out by health-care personnel in data collection sites should generally be avoided or minimized. Indeed, a major principle in the conduct of field studies is to minimize the work needed from cooperating clinical sites wherever feasible, since their continuing willingness to support the project is essential to its success. If CRFs are required, they should be as simple as possible, both to reduce errors and the level of effort required to complete them. An alternative is for study staff to directly abstract information from medical records, either on-site or at the study office. However, this is an arduous and labor-intensive process, particularly when the records are on paper rather than electronic. The format of paper medical records is not standardized, and the quality and completeness of the information varies greatly. If the abstraction is to be done at the study headquarters instead of the clinical sites, copies of the medical records must be requested. The balance between this additional step and the logistical complications and expense of travelling to the individual sites varies depending on the number and location of the sites. In either instance, as with CRFs, the abstraction forms should be designed to record only essential information, although it can be difficult to resist the temptation to include more rather than less.

Self-administered questionnaires are commonly used in large field studies, especially follow-up studies, which may involve tens of thousands of subjects.[25–27] Because there is no interviewer to control and standardize the process of filling out a questionnaire, it must be designed to lead the subject very clearly through the steps. As much as possible, numerical answers or check boxes are to be preferred over free text, which is not readily amenable to quantitative analysis. If the questionnaire is being completed online, it can incorporate branching that leads the respondent through the questions and sections in logical order contingent on their responses at each point, as well as error checks that direct the respondent to fill in missed questions or that flag out-of-range answers. Frequently, standardized validated instruments are incorporated into study questionnaires to obtain information on psychosocial factors,[28–32] dietary history,[33,34] physical activity,[35] etc. It is always good practice to take advantage of previously developed (and if possible, validated) instruments when these are available, since they have been shown to work in the field and produce results that can be more readily compared with other research. Specifically with regard to medication histories, standardized instruments are rare and the questions will likely have to be tailored to the specific needs of the project. As always, it remains desirable to keep a questionnaire as brief as possible to maximize its acceptability to subjects and hence the completion rate, while still covering the needed information.

The principles guiding the design of interview questionnaires (also see Chapter 41) are in many ways similar to those covering self-administered questionnaires, but there are also differences. A major difference is that a well-trained interviewer can be relied upon to lead the subject through the questions, which can allow for a less rigid structure that permits probing for additional information. There is a spectrum of views about how much latitude the interviewer should be given in going through the questions, which is reflected in questionnaire design choices. At one end is a highly scripted interview that is simply administered verbatim. Although this approach is very standardized, which is desirable, and also minimizes the training and educational requirements for interviewers (and hence staff costs), it is more difficult to establish rapport with the subject and tight scripting does not easily accommodate probing, which may be essential to obtaining certain types of information. The other end of the spectrum is a very open-ended approach, which has the important drawback of reducing consistency, and is more often used in qualitative research.[36] Perhaps most effective is a middle ground that retains structure and scripts for at least some of the questions and statements while still allowing interviewers latitude to probe and develop a style that is successful for them. This requires interviewers with a higher level of education than does a tightly scripted approach, for example nurses, and also places considerable emphasis on training and ongoing monitoring to maintain consistency. Regardless of the degree of scripting, simple questions with unequivocal numerical or yes/no answers are desirable, and standardized instruments—designed for interview rather than self-completion—are often used where applicable.

Medication histories in case–control studies are frequently obtained by interview (also see Chapter 41), and the development of appropriate questionnaires to elicit this information is of particular importance in pharmacoepidemiology. Questions on medication use can range from open-ended "did you take any drugs in the last month?", to asking about the use of specific medications of interest by name, and even showing handcards with drug names. An approach that has been proven to be effective in many studies, of which some examples are cited here,[7,9,21,23,37–56] is to ask systematically about drug use for specific indications—"did you take anything for pain in the last month?" A methodological study in 1986 that evaluated these different approaches sequentially[57] found that 0–45% of the use of a number of drugs was identified by an open-ended question, with a structured list of reasons adding 35–81%, and finally a specific question by name adding 19–39%. From this it can be inferred that asking about use by indication would cover 61–81% of the total. If only a few drugs are of particular interest, it is desirable to ask about them by name, especially if it is important to obtain

a detailed medication name, which for nonprescription drugs often has implications for dose as well as ingredients (e.g., Tylenol Allergy Sinus Daytime vs. Tylenol Allergy Sinus Nighttime). Requesting subjects to check the medication packages, or using product photographs during in-person interviews, have been shown to be helpful.[52,58] Another memory aid which is commonly used is a diary of life events.

Obtaining the names of medications taken is only half the battle, since details of use also need to be recorded. An effective approach in an interview situation is to first build a list of all the medications, and then go back to obtain the details.[52] The level of detail is determined by the research question, but also by practical considerations of what a subject can be expected to report with reasonable accuracy. Precise information on timing and amount of use is generally relevant only for the evaluation of acute effects, where a recent exposure period is of primary interest, for example, use of an antibiotic on days X, Y, and Z in the previous 2 weeks in relation to the development of agranulocytosis.[43] Studies of long-term effects generally require information on substantial use with less fine detail, which is likely to be well-remembered even if in the more distant past, for example use of an antihypertensive agent for at least 1 year in relation to lung cancer.[59]

Setup

Three key aspects of the setup of a field study involve the data collection network, the study staff, and the computing infrastructure (including the main database).

Data collection network

The particulars of the network are determined by the study population, sample size, and any specific considerations that require data from particular countries. These decisions need to be made in developing the protocol. The setup tasks are first to identify and then to recruit the needed institutions or practices. In some instances it may be possible to conduct a field study in a single center, but in practice the sample size requirements of pharmacoepidemiologic studies usually call for data collection in multiple centers. Any multicenter study will need some type of coordinating center to maintain the overall operation and ensure consistency of data collection. Generally, the coordinating center is under the direction of the overall project principal investigator, and has the responsibility for setting up the study network. If the study will be conducted in multiple countries, it will be necessary to have at least one co-investigator in each country who has knowledge of local conditions and connections with the appropriate institutions, and who can be responsible for setting up and running the project there, under the supervision of the coordinating center. A key factor in the success or failure of a large multicenter study, especially if it is conducted in multiple countries, is the level of engagement of the co-investigators. Thus, building a good collaboration is the essential first step in getting the study off the ground.[7]

In case–control studies in particular, cases are generally identified through the health-care system, most often relevant hospital departments, but occasionally doctors' offices. Recruiting the institutions is likely to be a large challenge, especially with modern privacy regulations and other human subjects considerations.[60–62] Large multicenter studies require applications to many IRBs, a process that can take months to complete. These applications require a local investigator in each institution, who will also be responsible for dealing with problems that arise, with the essential support of the coordinating center to ensure that the burden is minimized. In large hospital networks, where the individual investigators cannot practically be included as co-investigators in the core study team, active involvement with the participating physicians by the coordinating center is especially important. The participation of local co-investigators requires their commitment to the topic and minimization of their effort in the setup and operation of the study. Often the setup is a larger burden than the actual ongoing operation of the network when the study is up and running.

Study staff

As with other clinical studies that involve data collection, including randomized trials, an

experienced staff is an important component of success. One of the key positions is the study coordinator, who is responsible for the day-to-day running of the data collection network. Typical responsibilities for the coordinator include training and supervising field and office staff, monitoring the progress of subject enrollment and data collection, monitoring or performing quality control of the data, communication with co-investigators in multicenter studies, assisting with IRB applications, and providing information for data analyses. Because a clinical background is often needed to understand the data collection needs and interaction with the health-care system required for the enrollment of subjects, coordinators are frequently nurses. Similarly, medical knowledge is often necessary for personnel with direct responsibility for subject enrollment and data collection (especially if by interview, where sensitive medical questions are often asked), and these positions may be best filled by nurses, although sometimes non-clinical research assistants are appropriate. It is advisable and sometimes a requirement because of privacy or other institutional regulations for data collection personnel in specific hospitals to be employees of those institutions. Other central staff may include research assistants, coders, and research pharmacists. Another key position, although not unique to field studies, is that of data analyst, since the principal investigator and co-investigators generally do not fill this role. A minimum of master's level training is normally required. Studies that involve data collection frequently require the input of experienced clinicians to ensure that subjects meet inclusion criteria. In such situations it is often appropriate to engage the appropriate specialists as consultants to the study team.

Computing infrastructure

In modern field studies, the importance of the computer infrastructure can scarcely be overstated. It behooves any group that is continually engaged in the conduct of these studies to build a good programming and computer operations department. There are three main components to the infrastructure: (i) logs to track subject enrollment and data collection; (ii) software for data collection and management, which may include among other modules, data entry software for computerized interviews, online questionnaires, scanning software for paper forms, quality control and coding software, and a system for producing automated letters to subjects and providers; and (iii) the study database itself. The process of developing these components usually requires several months, although it can be shortened somewhat by adapting existing systems. Field testing is generally recommended. While a substantial effort may be needed to build the computer infrastructure, this yields many benefits, including more efficient study operations, fewer errors in the data, and a streamlined process of preparing data for analysis. The time spent during the early phase is more than compensated for by savings later on.

While the setup process is non-trivial for field studies, the time required can be reduced if existing staff, collaborator networks, or components of the computer infrastructure can be utilized, but, even then, it is generally not reasonable to expect that a study can be up and running within a month or two. A realistic timeline for setup will lead to better results during the conduct of the study, and ultimately make for better relations between investigator and sponsor.

Pilot phase

During the set-up phase, it is generally advisable to test the methods by initiating data collection in a few sites. This pilot period allows adjustment of study procedures before full-scale operations begin. For example, it may be that the initial inclusion criteria do not yield the expected number of subjects, and need to be relaxed (after determination that this will not jeopardize validity). The originally planned subject identification and enrollment procedures in the study sites may prove to be infeasible or at least overly inefficient, with pilot experience suggesting modifications that substantially improve the process. Experience with the data collection instruments may point the way to improvements that increase the likelihood that the desired information will be obtained. In all of these areas, obtaining initial experience on a manageable scale is generally a prudent approach to finalizing the

setup of a study for routine data collection. A pilot phase may not be necessary if the method is very well proven, but since field studies are usually mounted to investigate questions that cannot be answered with existing data, the operational problems are frequently new, requiring a fresh approach where the feasibility must first be demonstrated before implementing on a full-scale basis.

Conduct of the study

Most pharmacoepidemiologic field studies are multicenter in nature, giving rise to some general operational principles that underlie their conduct:

• *Communication*—the imperative of maintaining good communication between the coordinating center and field operations and among co-investigators cannot be stressed too highly. Communication is necessary to understand what is happening at the study sites, to maintain consistency in the conduct of the study throughout the network, and to keep a geographically separated team of investigators fully informed of problems and involved in the solutions. Good communication requires substantial effort but it always reaps dividends and should be made a very high priority.

• *The study team should actively conduct the study*—particularly in subject enrollment, but also in data collection, a passive approach that relies on the good will and efforts of individuals who are not formally part of the study team is likely to lead to recruitment difficulties and substandard data. The primary commitment of clinicians and their staffs is to provide medical care, and as much as possible, a study should rely on their granting access to patients and information and nothing more. Voluntary referrals of subjects are likely to be biased and in low numbers. Expectations that clinical personnel will actively assist with data collection, whether through filling out case report forms or interacting with subjects, are generally unrealistic. It has been said many times that "you can't pay them enough" to pursue the study with the degree of attention expected of staff employed to conduct the project. Observational field studies are to be contrasted in this regard with clinical trials, which in addition to larger budgets devoted to activities at individual sites, often have the commitment of participating physicians who are seeking more effective therapies for their patients. Further, even if the subjects recruited are atypical (which has been shown repeatedly), they are then randomized, so internal validity is protected; selection bias cannot be introduced by incomplete enrollment. However, there is recent evidence that even many randomized trials are unsuccessful in recruiting enough patients.[63] Even without an extensive budget for the sites, a highly active approach to the conduct of field studies is more expensive, but it is essential to producing valid, informative results in a timely manner.

• *Do not cut corners to save costs or time*—there may be pressure from sponsors, especially commercial ones, to accelerate the timetable and cut costs. Field studies are usually commissioned by industry only when studies of existing data, which can be conducted rapidly and at relatively modest cost, are infeasible. Thus, sponsors need to be aware of the different realities that apply to field study operations. While it is important to be parsimonious in the use of resources and mindful of the calendar as a study proceeds, it is equally important to be realistic about what is needed to conduct a high-quality investigation that will achieve its goals. A substandard study conducted rapidly without sufficient financial resources serves no one's interests.

Subject enrollment

There are two key goals in the enrollment phase of all field studies: to reach the targeted number of subjects meeting the inclusion criteria so as to meet the sample size requirements, and to maximize the participation rate in order to reduce the possibility of selection bias. There are numerous approaches to achieving these goals. Follow-up studies, which tend to be very large, may enroll subjects through the use of magazine subscriber lists,[64] registries of professionals,[25] and advertisements.[27] The numbers of subjects in case–control studies are usually much smaller, a few hundred or, at most, a few thousand cases and controls, but the enrollment of cases of rare conditions that are often of interest in pharmacoepidemiology (e.g., blood dyscrasias) may require large and even international networks to obtain the needed numbers.[7,9,65,66]

Cases are most often identified through contact with the health-care system, and the principle of active ascertainment, which has already been mentioned and involves members of the study staff taking the initiative to identify cases, is the most important defense against selection bias during the enrollment process. Specific approaches include regular telephone contact with the relevant sites,[37] or visits by study staff to identify subjects and approach them for participation.[44] The rapid turnaround and reliability of email has opened up further possibilities for efficiently communicating with sites. A complicating factor is privacy regulations, which in the US now prohibit outside personnel from examining medical records and approaching patients without prior permission or a waiver of authorization.[67,68] On the other hand, medical personnel can examine records with a Health Insurance Portability and Accountability Act (HIPAA) preparatory to research waiver,[69] and health-care providers can then invite potentially eligible subjects to participate, either in person or by letter. The latter process can be simplified for the sites by developing a computerized system to generate letters of invitation. Patients approached in this way can either indicate their agreement to have contact information forwarded to the study staff (opt-in) or indicate by mail or phone that they do not wish to participate, in which case the information is not forwarded (opt-out). With the opt-out approach, the contact information is forwarded to the study staff if the subject has not opted out within a specified time period. In either instance, once the contact information has been sent to the study office, actual enrollment, including informed consent, is controlled by study personnel and not reliant on individual sites.

The different recruitment mechanisms require IRB approval for the individual study sites. For reasons related to study validity and practicality, opt-out is greatly to be preferred to opt-in, because it usually results in higher participation rates of eligible subjects. However, many IRBs will only approve an opt-in approach, which can sometimes make a study infeasible. Variations on the opt-out approach exist, for example, recruiting Medicare recipients in the US through the Centers for Medicare and Medicaid Services (CMS).[70] This has the further advantage of initially identifying subjects from a comprehensive list, which is less subject to bias. Other list-based recruitment approaches, such as identifying potential subjects through cancer registries,[71,72] may occasionally be feasible, and preferable to working with clinical sites because they do not require the step of medical personnel identifying subjects to approach.

More problematic enrollment approaches include those that have clinical personnel recruit and consent subjects, or some form of advertising, for example by internet, flyers, newspapers, or through patient organizations. The former is logistically cumbersome, requires an impractical level of effort from the sites, and is likely to lead to low and biased recruitment. It is important here to distinguish the situation of an observational study from a clinical trial where eligible subjects are randomized. In the trial setting, low recruitment may make it difficult to obtain an adequate sample size, but it does not lead to bias. Advertising is a much more passive approach, and the potential for biased enrollment in an observational study based on who sees the ads and is interested enough to volunteer for a study is obvious.

The optimal enrollment situation is one that is controlled by study staff as early in the process as possible. Once a potentially eligible subject is identified, assiduous efforts must be made to contact her or him, determine final eligibility through some sort of screening process (e.g., a brief screening interview or questionnaire), and obtain consent to participate. Depending on the nature of the data collection, this can be done by phone, in person, or by mail (interview studies may not require that informed consent be documented with the subject's signature). Meeting the general goal of enrolling as many identified eligible subjects as possible often requires multiple contacts by telephone, email, or regular mail.

In certain instances a more elaborate process for determining the eligibility of cases is required to ensure that uniform criteria have been met. This is particularly important in studies of rare conditions for which the diagnoses are not straightforward and subjects are enrolled from multiple countries

where health-care practices may be different. For example, in two multicenter studies of Stevens–Johnson syndrome and toxic epidermal necrolysis, it was deemed necessary to have a committee of dermatologists review photographs as well as relevant clinical and pathological information to confirm the diagnoses.[8,10] The review committees met twice a year during the course of the studies. A similar approach was taken for agranulocytosis and aplastic anemia in the IAAAS,[6,7] where actual biopsy specimens were obtained for review for more than 95% of the cases; another example of the use of a review committee was in an international study of acute anaphylaxis in relation to drug exposures.[22] This general approach assures the most consistent case definitions, but requires considerable resources to assemble all of the relevant materials and cover the costs of the review sessions. Thus, it should only be undertaken when a more localized confirmation process in individual centers is judged to be unlikely to yield comparable diagnostic standards.

The enrollment of controls in case–control studies may be done through health-care providers (e.g., other hospitalized patients), which generally requires permission from providers to approach potential subjects for enrollment. This, of course, risks selection bias if the illnesses which result in the medical attention that qualifies the controls are related to the exposure of interest, and also from incomplete enrollment. Appropriate specification of diagnoses for hospital controls is a key design issue that requires considerable attention in developing protocols for hospital-based case–control studies. "Population" or "community" controls can be identified by means of random digit dialing,[73] lists of licensed drivers,[74] Medicare rolls,[75] or municipal census lists.[21,47] With the latter approaches, identified potential controls can then be approached for participation directly by study staff, except in the instance of Medicare, where subjects are first approached by CMS, with their contact information given to investigators after an opt-out period has expired. Another possibility is nominations by cases (e.g., relatives, friends), but this may have considerable potential for bias.[7] Advertising is subject to the concerns noted above. Maximum effort by study personnel to contact and enroll targeted hospital or community controls is crucial for validity as well as efficiency.

Data collection

Once subjects are enrolled, the primary goal of the data collection phase of a field study is to obtain the specified data with as much completeness and consistency as possible. With self-administered data collection procedures, including mailed or internet questionnaires, study staff need to follow up with subjects who do not complete them in a timely manner. The follow-up process can involve repeat mailings and telephone or email contact as appropriate. To ensure maximum participation in this phase of a study, it may be advisable to offer the option of an interview for subjects who will not complete the questionnaires on their own.

Interviews (see also Chapter 41) are often the primary source of data in pharmacoepidemiologic field studies, and have the advantage over self-administered questionnaires of being a more consistent and controlled process, albeit more costly. The importance of careful training of interviewers to maximize the consistency and quality of the data they collect has already been discussed. Interviews can either be conducted in person (in the hospital, at a subject's home, or at some other site) or by telephone. The nature of the study often dictates the interview setting. In a case–control study where subjects are enrolled while in hospital, an in-person interview is to be preferred. On the other hand, when subjects are identified after they have gone home, or community controls are used, a telephone interview is more feasible. It is important for comparability that the interview setting be as similar as possible for cases and controls.

Regardless of the type of interview, modern field studies generally involve the use of computerized data entry systems; this technology is referred to as computer-assisted telephone interview (CATI)[76] and computer-assisted personal interview (CAPI).[77] CATI is used when interviews are conducted by telephone from a call center, research office, or even the interviewer's home. CAPI involves the use of laptop computers to conduct interviews in person. In either instance, the software operates

similarly, providing the questions for the interviewer to ask and checkboxes or entry fields to record the answers. In CAPI, where the laptop is not connected to a central server, all the necessary software is resident on the computer and the data are saved locally and uploaded later. Because computerized interviews save a separate data entry step that would be needed with paper questionnaires, they are cost-effective and reduce entry errors. Sophisticated branching of questions can be readily accommodated, leading the interviewer and subject through what may be a complex set of questions. Another advantage of computerized interviews is that automatic coding during the interview, for example for drug product names, can be built into the software, increasing accuracy and again saving a data processing step.

If data are obtained from medical records, these must be requested (with a signed release from the subject, except in the unusual circumstance where an IRB has waived the requirement for informed consent) and abstracted by study staff, often a labor-intensive process that may require repeat requests to providers for the information. In some instances with relatively small data collection networks, it is feasible for study staff to actually visit the sites and abstract the information there, although this has become uncommon because of privacy regulations.[67,68]

Some field studies require the collection of biological samples, for example blood, urine, or tissue (such as biopsy material). With today's focus on genetics, DNA samples are increasingly sought. DNA can be obtained either from blood samples or from cheek cells collected through buccal swabs or saliva samples, which can be done by subjects at home. A drawback to cheek cell samples is the restricted amount and reduced quality of DNA compared to that obtained from whole blood, but it is sufficient for many purposes.[24,78] Urine and stool specimens can also be collected at home[79] or at a study site. If blood samples are needed, these must be drawn by a professional, either at the subject's home or at some sort of collection site, which complicates logistics and adds to costs. In addition to procedures related to the sample collection itself, freezers and other storage facilities must be maintained, and relationships with appropriate laboratories developed for conducting the relevant tests.

Data management

An important aspect of the data management process is coding and quality control (QC). Many data items, including drug product names, medical conditions, procedures, occupation, and so on, may be collected as text that must be numerically coded for analysis. Interview software can provide a head start on this process by linking with computerized dictionaries that will automatically code data entries that match up with dictionary entries, but there is always a need for at least some coding after data collection because of misspelled entries or those that require some interpretation. In some studies, all coding is done after the data have been collected.

Appropriate coding can make a major contribution to efficient analyses. Diagnoses are frequently coded with International Classification of Diseases (ICD) codes.[80] For drug names, there are various dictionaries available, including among others, the WHO Drug Dictionary (WHO-DD),[81] Iowa Drug Information Service (IDIS) Drug Vocabulary and Thesaurus,[82] and the Slone Drug Dictionary.[83] The WHO dictionary includes drug names linked to the Anatomical Therapeutic Chemical (ATC) Classification System[84] while the IDIS and Slone dictionaries utilize a modified version of the American Hospital Formulary Service (AHFS) Pharmacologic Therapeutic Classification.[85] The classification systems facilitate the identification and grouping of agents with similar pharmacologic properties and therapeutic uses. These dictionaries have been used in various pharmacoepidemiologic studies and have varied content. While no single dictionary is comprehensive for all prescription and non-prescription medications, each of them is well suited for the coding of most drug names that have been encountered in Europe and the United States. The IDIS is limited to mostly US prescription drug names linked to their therapeutic class. The Slone dictionary contains US and non-US drugs, and in addition it includes supplements and other non-prescription products; entries are cross-linked with their individual drug, vitamin, and herbal ingredi-

ents, which readily allows for grouping products that contain the same ingredient (e.g., acetaminophen) and drug class (e.g., selective serotonin reuptake inhibitor antidepressants). The WHO dictionary contains primarily European drug names and their ingredients, and has recently added a herbal dictionary based on the Herbal ATC system (HATC).[86]

The QC process is an important step in ensuring that the data are consistent and as free from errors as possible. The first line of defense with computerized interviews or online self-administered questionnaires is built-in checks in the software to not accept entry of inappropriate answers, for example out of range ages, impossible height and weight values. Another common feature that maximizes completeness of the information is to require that certain questions be answered before the interview can proceed. Good practice for interviewers is to review the questionnaires (whether computerized or on paper) immediately after completion to clean up inconsistencies in filling them out. At the central office there are usually two levels of QC. Another series of automatic checks built into the software for uploading interview and self-administered data to the main database flags inappropriate entries for correction. There should also be some level of manual QC of the data, which can range from a complete review of all items for each subject in small studies to review of key items such as medication histories or random spot checking the data for individual subjects in very-high-volume studies where individual checking of all data is not feasible. Errors and inconsistencies identified at this stage can be corrected by contacting the interviewer and sometimes even the subject if necessary. After all QC checks have been completed, the data can be released for use in analyses.

The process of uploading data to central databases is usually fairly automated in modern field studies. Computerized interviews and online self-administered questionnaires can be transferred directly, with some built-in QC checks as described above. Paper questionnaires are designed to be scanned whenever possible, although text fields may require manual entry. Abstraction of medical records is ideally accomplished with a computerized form that again allows direct uploading to a central database. When computer-assisted data collection processes such as interviews and abstraction are done locally rather than from a central office, automated uploading is still possible via secure internet connection. Entry of laboratory test results and other similar data items may require a manual process, which can be greatly facilitated by appropriately designed software.

A common theme throughout this discussion of the conduct of field studies is that a well-integrated computing infrastructure plays a major role in many aspects of the process, from subject enrollment, through data collection, to the final processing of the data prior to analysis. An important counterpoint, however, is that while computerization provides many advantages of efficiency and error reduction, there is no substitute for the human element in the collection of high-quality data for valid and informative study results.

Analysis

Most analytical issues are not unique to field studies, and will not be covered here. A sometimes underappreciated step that is worth mentioning involves the creation from the raw data of analytical files that contain the variables on exposures, outcomes, and covariates needed for a particular analysis. The goal in the data collection phase of a study is to obtain information on all items that may be needed to address all the research questions. The format of the data should be designed to facilitate accurate and complete collection, and may not be appropriate for analysis without some transformation. For example, complex medication histories may need to be simplified into analytical variables by combining information from the use of several products, calculating total dosage ingested over a specified time period, and defining the temporal sequence of drug use and clinical events. The latter process may require comparing information on the timing of drug use and the clinical course of a medical condition that has been obtained using a common reference point that is consistent for all subjects (e.g., the day of hospital admission). This "preanalysis" creation of analytical files requires first a conceptual process to define the variables

needed for the analysis. The actual preparation of a file can involve substantial computing from the raw data. Whether in a field study or a database study, once an analytical file is prepared, the epidemiologic analyses, while defined by the data items and specific questions addressed, are similar.

The future

As databases (see Part III Section B) and other ongoing resources such as registries (see Chapter 21) become more computerized and better adapted for research, there will be less call for field studies (although a registry often emerges from a special type of field study). However, it is not conceivable that the day will come where all pharmacoepidemiologic research questions can be answered with already-collected data. There will always be situations that require more detailed information about the disease under study than is available from preexisting data, information about OTC medication use, details of habits that are not routinely recorded, or information on factors that can only be provided by subjects, such as quality of life. In these instances, field studies will be needed. It is to be hoped that investigators and organizations with the capacity to mount such studies continue to be available, so that the conduct of field studies does not become a lost art in the practice of pharmacoepidemiology.

Acknowledgement

I wish to thank my colleague, Lynn Rosenberg, for her many helpful comments.

References

1. Graham DJ, Campen D, Hui R, Spence M, Cheetham C, Levy G, *et al.* Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. *Lancet* 2005; **365**: 475–81.
2. Bastuji-Garin S, Rzany B, Stern RS, Shear NH, Naldi L, Roujeau JC. Clinical classification of cases of toxic epidermal necrolysis, Stevens-Johnson syndrome, and erythema multiforme. *Arch Dermatol* 1993; **129**: 92–6.
3. Strom BL. Data validity issues in using claims data. *Pharmacoepidemiol Drug Saf* 2001; **10**: 389–92.
4. Spriet-Pourra C, Auriche M. *Drug Withdrawal from Sale: An Analysis of the Phenomenon and Its Implications*. Richmond, UK: PJB Publications Ltd, 1988.
5. Fung M, Thornton A, Mybeck K, Wu JH, Hornbuckle K, Muniz E. Evaluation of the characteristics of safety withdrawal of prescription drugs from worldwide pharmaceutical markets—1960 to 1999. *Drug Inf J* 2001; **35**: 293–317.
6. Anon. The design of a study of the drug etiology of agranulocytosis and aplastic anemia. A report from the International Agranulocytosis and Aplastic Anemia Study. *Eur J Clin Pharmacol* 1983; **24**: 833–6.
7. Kaufman DW, Kelly JP, Levy M, Shapiro S. *The Drug Etiology of Agranulocytosis and Aplastic Anemia: The International Agranulocytosis and Aplastic Anemia Study*. New York: Oxford University Press, 1991.
8. Kelly JP, Auquier A, Rzany B, Naldi L, Bastuji-Garin S, Correia O, *et al.* An international collaborative case-control study of severe cutaneous adverse reactions (SCAR). Design and methods. *J Clin Epidemiol* 1995; **48**: 1099–108.
9. Roujeau JC, Kelly JP, Naldi L, Rzany B, Stern RS, Anderson T, *et al.* Medication use and the risk of Stevens-Johnson syndrome or toxic epidermal necrolysis. *N Engl J Med* 1995; **333**: 1600–7.
10. Mockenhaupt M, Viboud C, Dunant A, Naldi L, Halevy S, Bouwes Bavinck JN, *et al.* Stevens-Johnson syndrome and toxic epidermal necrolysis: assessment of medication risks with emphasis on recently marketed drugs. The EuroSCAR-study. *J Invest Dermatol* 2008; **128**: 35–44.
11. Rosenberg L, Palmer JR, Adams-Campbell LL. Postmenopausal female hormone use and venous thromboembolic disease in black women. *Am J Obstet Gynecol* 1997; **177**: 1275.
12. Grodstein F, Manson JE, Colditz GA, Willett WC, Speizer FE, Stampfer MJ. A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann Intern Med* 2000; **133**: 933–41.
13. Chen WY, Manson JE, Hankinson SE, Rosner B, Holmes MD, Willett WC, *et al.* Unopposed estrogen therapy and the risk of invasive breast cancer. *Arch Intern Med* 2006; **166**: 1027–32.
14. Lee JE, Hankinson SE, Cho E. Reproductive factors and risk of renal cell cancer: the Nurses' Health Study. *Am J Epidemiol* 2009; **169**: 1243–50.

15. Kaufman DW, Kelly JP, Jurgelon JM, Anderson T, Issaragrisil S, Wiholm BE, *et al.* Drugs in the aetiology of agranulocytosis and aplastic anaemia. *Eur J Haematol* 1996; **60** (Suppl): 23–30.
16. Rzany B, Correia O, Kelly JP, Naldi L, Auquier A, Stern R. Risk of Stevens-Johnson syndrome and toxic epidermal necrolysis during first weeks of antiepileptic therapy: a case-control study. Study Group of the International Case Control Study on Severe Cutaneous Adverse Reactions. *Lancet* 1999; **353**: 2190–4.
17. Kaufman DW. Epidemiologic approaches to the study of toxic epidermal necrolysis. *J Invest Dermatol* 1994; **102**: 31S–3S.
18. Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol* 2007; **17**: 643–53.
19. McLaughlin JK, Lipworth L, Chow WH, Blot WJ. Analgesic use and chronic renal failure: a critical review of the epidemiologic literature. *Kidney Int* 1998; **54**: 679–86.
20. Rosamond WD, Chambless LE, Folsom AR, Cooper LS, Conwill DE, Clegg L, *et al.* Trends in the incidence of myocardial infarction and in mortality due to coronary heart disease, 1987 to 1994. *N Engl J Med* 1998; **339**: 861–7.
21. Kaufman DW, Kelly JP, Sheehan JE, Laszlo A, Wiholm BE, Alfredsson L, *et al.* Nonsteroidal antiinflammatory drug use in relation to major upper gastrointestinal bleeding. *Clin Pharmacol Ther* 1993; **53**: 485–94.
22. Kaufman DW, Kelly JP, for the International Collaborative Study of Severe Anaphylaxis. Risk of anaphylaxis in a hospital population in relation to the use of various drugs: an international study. *Pharmacoepidemiol Drug Safety* 2003; **12**: 195–202.
23. Sakamoto C, Sugano K, Ota S, Sakaki N, Takahashi S, Yoshida Y, *et al.* Case-control study on the association of upper gastrointestinal bleeding and nonsteroidal anti-inflammatory drugs in Japan. *Eur J Clin Pharmacol* 2006; **62**: 765–72.
24. Cozier YC, Palmer JR, Rosenberg L. Comparison of methods for collection of DNA samples by mail in the Black Women's Health Study. *Ann Epidemiol* 2004; **14**: 117–22.
25. Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses' health study. *Am J Nurs* 1978; **78**: 1039–40.
26. Rosenberg L, Adams-Campbell L, Palmer JR. The Black Women's Health Study: a follow-up study for causes and preventions of illness. *J Am Med Womens Assoc* 1995; **50**: 56–8.
27. Anon. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials* 1998; **19**: 61–109.
28. Hewitt P, Norton G. The Beck Anxiety Inventory: a psychometric analysis. *Psychol Assess* 1993; **5**: 408–12.
29. Turvey CL, Wallace RB, Herzog R. A revised CES-D measure of depressive symptoms and a DSM-based measure of major depressive episodes in the elderly. *Int Psychogeriatr* 1999; **11**: 139–48.
30. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001; **33**: 337–43.
31. Kazis LE, Miller DR, Skinner KM, Lee A, Ren XS, Clark JA, *et al.* Patient-reported measures of health: The Veterans Health Study. *J Ambul Care Manage* 2004; **27**: 70–83.
32. Williams CD, Taylor TR, Makambi K, Harrell J, Palmer JR, Rosenberg L, *et al.* CES-D four-factor structure is confirmed, but not invariant, in a large cohort of African American women. *Psychiatry Res* 2007; **150**: 173–80.
33. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, *et al.* Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985; **122**: 51–65.
34. Kumanyika SK, Mauger D, Mitchell DC, Phillips B, Smiciklas-Wright H, Palmer JR. Relative validity of food frequency questionnaire nutrient estimates in the Black Women's Health Study. *Ann Epidemiol* 2003; **13**: 111–8.
35. Paffenbarger RS, Jr., Blair SN, Lee IM, Hyde RT. Measurement of physical activity to assess health effects in free-living populations. *Med Sci Sports Exerc* 1993; **25**: 60–70.
36. Mack N, Woodsong C, MacQueen KM, Guest G, Namey E. *Qualitative Research Methods: A Data Collector's Field Guide*. Research Triangle Park, NC: Family Health International, 2005.
37. Shapiro S, Slone D, Rosenberg L, Kaufman DW, Stolley PD, Miettinen OS. Oral-contraceptive use in relation to myocardial infarction. *Lancet* 1979; **1**: 743–7.
38. Rosenberg L, Slone D, Shapiro S, Kaufman DW, Stolley PD, Miettinen OS. Noncontraceptive estrogens and myocardial infarction in young women. *JAMA* 1980; **244**: 339–42.
39. Shapiro S, Kaufman DW, Slone D, Rosenberg L, Miettinen OS, Stolley PD, *et al.* Recent and past use of conjugated estrogens in relation to adenocarcinoma of the endometrium. *N Engl J Med* 1980; **303**: 485–9.

40. Kaufman DW, Shapiro S, Slone D, Rosenberg L, Miettinen OS, Stolley PD, *et al.* Decreased risk of endometrial cancer among oral-contraceptive users. *N Engl J Med* 1980; **303**: 1045–7.
41. Rosenberg L, Shapiro S, Slone D, Kaufman DW, Helmrich SP, Miettinen OS, *et al.* Epithelial ovarian cancer and combination oral contraceptives. *JAMA* 1982; **247**: 3210–2.
42. Kaufman DW, Miller DR, Rosenberg L, Helmrich SP, Stolley P, Schottenfeld D, *et al.* Noncontraceptive estrogen use and the risk of breast cancer. *JAMA* 1984; **252**: 63–7.
43. International Agranulocytosis and Aplastic Anemia Study. Anti infective drug use in relation to the risk of agranulocytosis and aplastic anemia. *Arch Intern Med* 1989; **149**: 1036–40.
44. Kaufman DW, Werler MM, Palmer JR, Rosenberg L, Stolley PD, Warshauer ME, *et al.* Diazepam use in relation to breast cancer: results from two case-control studies. *Am J Epidemiol* 1990; **131**: 483–90.
45. Laporte JR, Carne X, Vidal X, Moreno V, Juan J. Upper gastrointestinal bleeding in relation to previous use of analgesics and non-steroidal anti-inflammatory drugs. Catalan Countries Study on Upper Gastrointestinal Bleeding. *Lancet* 1991; **337**: 85–9.
46. Werler MM, Mitchell AA, Shapiro S. First trimester maternal medication use in relation to gastroschisis. *Teratology* 1992; **45**: 361–7.
47. Rosenberg L, Palmer JR, Shapiro S. A case-control study of myocardial infarction in relation to use of estrogen supplements. *Am J Epidemiol* 1993; **137**: 54–63.
48. Kelly JP, Kaufman DW, Jurgelon JM, Sheehan J, Koff RS, Shapiro S. Risk of aspirin-associated major upper-gastrointestinal bleeding with enteric-coated or buffered product. *Lancet* 1996; **348**: 1413–6.
49. Issaragrisil S, Kaufman DW, Anderson T, Chansung K, Thamprasit T, Sirijirachai J, *et al.* Low drug attributability of aplastic anemia in Thailand. The Aplastic Anemia Study Group. *Blood* 1997; **89**: 4034–9.
50. Rosenberg L, Louik C, Shapiro S. Nonsteroidal anti-inflammatory drug use and reduced risk of large bowel carcinoma. *Cancer* 1998; **82**: 2326–33.
51. 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**: 961–8.
52. 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**: 337–44.
53. Rosenberg L, Rao RS, Palmer JR. A case-control study of acetaminophen use in relation to the risk of first myocardial infarction in men. *Pharmacoepidemiol Drug Saf* 2003; **12**: 459–65.
54. Louik C, Lin AE, Werler MM, Hernandez-Diaz S, Mitchell AA. First-trimester use of selective serotonin-reuptake inhibitors and the risk of birth defects. *N Engl J Med* 2007; **356**: 2675–83.
55. Coogan PF, Smith J, Rosenberg L. Statin use and risk of colorectal cancer. *J Natl Cancer Inst* 2007; **99**: 32–40.
56. Vidal X, Ibanez L, Vendrell L, Conforti A, Laporte JR. Risk of upper gastrointestinal bleeding and the degree of serotonin reuptake inhibition by antidepressants: a case-control study. *Drug Saf* 2008; **31**: 159–68.
57. Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.
58. Werler MM, Sheehan JE, Mitchell AA. Maternal medication use and risks of gastroschisis and small intestinal atresia. *Am J Epidemiol* 2002; **155**: 26–31.
59. Kaufman DW, Kelly JP, Rosenberg L, Stolley PD, Warshauer ME, Shapiro S. Hydralazine use in relation to cancers of the lung, colon, and rectum. *Eur J Clin Pharmacol* 1989; **36**: 259–64.
60. Hiatt RA. HIPAA: the end of epidemiology, or a new social contract? *Epidemiology* 2003; **14**: 637–9.
61. Vates JR, Hetrick JL, Lavin KL, Sharma GK, Wagner RL, Johnson JT. Protecting medical record information: start your research registries today. *Laryngoscope* 2005; **115**: 441–4.
62. Davidow AL, Katz D, Reves R, Bethel J, Ngong L. The challenge of multisite epidemiologic studies in diverse populations: design and implementation of a 22-site study of tuberculosis in foreign-born people. *Public Health Rep* 2009; **124**: 391–9.
63. IOM (Institute of Medicine). *A National Cancer Clinical Trials System for the 21st Century: Reinvigorating the NCI Cooperative Group Program*. Washington, DC: National Academies Press, 2010.
64. Rosenberg L, Palmer JR, Rao RS, Adams-Campbell LL. Risk factors for coronary heart disease in African American women. *Am J Epidemiol* 1999; **150**: 904–9.
65. Abenhaim L, Moride Y, Brenot F, Rich S, Benichou J, Kurz X, *et al.* Appetite-suppressant drugs and the risk of primary pulmonary hypertension. International Primary Pulmonary Hypertension Study Group. *N Engl J Med* 1996; **335**: 609–16.

66. Mary JY, Guiguet M, Baumelou E. Drug use and aplastic anaemia: the French experience. French Cooperative Group for the Epidemiological Study of Aplastic Anaemia. *Eur J Haematol Suppl* 1996; **60**: 35–41.
67. US Department of Health and Human Services. *Health Information Privacy*. http://www.hhs.gov/ocr/privacy/ (accessed May 2010).
68. US Department of Health and Human Services, National Institutes of Health. *HIPAA Privacy Rule Information for Researchers*. http://privacyruleandresearch.nih.gov/ (accessed May 2010).
69. US Department of Health and Human Services, National Institutes of Health. *How Can Covered Entities Use and Disclose Protected Health Information for Research and Comply with the Privacy Rule?* http://privacyruleandresearch.nih.gov/pr_08.asp (accessed May 2010).
70. Centers for Medicare and Medicaid Services. *CMS Standard Analytical Files*. http://www.cms.hhs.gov/IdentifiableDataFiles/02_StandardAnalyticalFiles.asp#TopOfPage (accessed June 2011).
71. Folger SG, Marchbanks PA, McDonald JA, Bernstein L, Ursin G, Berlin JA, *et al.* Risk of breast cancer associated with short-term use of oral contraceptives. *Cancer Causes Control* 2007; **18**: 189–98.
72. Mathes RW, Malone KE, Daling JR, Porter PL, Li CI. Relationship between histamine2-receptor antagonist medications and risk of invasive breast cancer. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 67–72.
73. Marchbanks PA, McDonald JA, Wilson HG, Folger SG, Mandel MG, Daling JR, *et al.* Oral contraceptives and the risk of breast cancer. *N Engl J Med* 2002; **346**: 2025–32.
74. Hall IJ, Moorman PG, Millikan RC, Newman B. Comparative analysis of breast cancer risk factors among African-American women and White women. *Am J Epidemiol* 2005; **161**: 40–51.
75. Lauderdale DS, Furner SE, Miles TP, Goldberg J. Epidemiologic uses of Medicare data. *Epidemiol Rev* 1993; **15**: 319–27.
76. Groves RM, Mathiowetz NA. Computer assisted telephone interviewing: effects on interviewers and respondents. *Public Opin Q* 1984; **48**: 356–69.
77. Groves RM, Mosher WD, Lepkowski JM, Kirgis NG. Planning and development of the continuous National Survey of Family Growth. *Vital Health Stat 1* 2009: 1–64.
78. Ruiz-Narvaez EA, Rosenberg L, Cozier YC, Cupples LA, Adams-Campbell LL, Palmer JR. Polymorphisms in the TOX3/LOC643714 locus and risk of breast cancer in African-American women. *Cancer Epidemiol Biomarkers Prev* 2010; **19**: 1320–7.
79. Kaufman DW, Kelly JP, Curhan GC, Anderson TE, Dretler SP, Preminger GM, *et al.* Oxalobacter formigenes may reduce the risk of calcium oxalate kidney stones. *J Am Soc Nephrol* 2008; **19**: 1197–203.
80. National Center for Health Statistics. *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*, 2009 http://www.cdc.gov/nchs/icd/icd9cm.htm (accessed May 2010).
81. Uppsala Monitoring Center. *World Health Organization (WHO) Drug Dictionary* http://www.umc-products.com (accessed May 2010).
82. Rumschlag D, Howes B. IDIS: Iowa Drug Information Service. *CD-Rom World* 1993; **6**: 80–3.
83. Kelley KE, Kelley TP, Kaufman DW, Mitchell AA. The Slone Drug Dictionary: A research driven pharmacoepidemiology tool. *Pharmacoepidemiol Drug Saf* 2003; **12**: S168–9.
84. WHO Collaborating Centre for Drug Statistics Methodology. *Guidelines for ATC Classification and DDD Assignment*, 2009.
85. American Society of Health-System Pharmacists . *AHFS Drug Information 2010*.
86. Uppsala Monitoring Centre. *Herbal ATC Index*, 2004.

CHAPTER 23
How Should One Perform Pharmacoepidemiologic Studies? Choosing Among the Available Alternatives

Brian L. Strom
Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

As discussed in the 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. This large sample size raises logistical challenges. Chapters 10 through 22 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 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 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 23.1). One can then examine the characteristics of the research question at hand, in order to choose the pharmacoepidemiologic approach best suited to addressing that question (see Table 23.2). The assessment and weights provided in this discussion and in the

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

Table 23.1 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)

Table 23.1 (*Continued*)

Pharmacoepidemiologic approach	Validity of exposure data	Validity of outcome data	Control of confounding	Inpatient drug exposure data	Outpatient diagnosis data	Loss to follow-up
Spontaneous reporting	+++	++	–	+++	+++	N/A
Health maintenance organizations / 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

*See the text of this chapter for descriptions of the column headings, and previous chapters for descriptions of the data resources.

Table 23.2 Characteristics of research questions and their impact on the choice of pharmacoepidemiologic data resources*

Pharmacoepidemiologic approach	Hypothesis generating†	Hypothesis strengthening‡	Hypothesis testing§	Study of benefits (versus risk)	Incidence rates desired	Low incidence outcome	Low prevalence exposure
Spontaneous reporting	++++	+	–	–	–	++++	++++
Health maintenance organizations / 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	+	+	++++	++++	++++	+	++++

Table 23.2 (*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 a new drug	Urgent question
Spontaneous reporting	–	+++	++++	+	+	++++	++++
Health maintenance organizations/ 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	++++	+++	++++	++++	+	++++	+

* See the text of this chapter for descriptions of the column headings, and previous chapters for descriptions of the data resources.
† Hypothesis-generating studies are studies designed to raise new questions about possible unexpected drug effects, whether adverse or beneficial.
‡ Hypothesis-strengthening studies are studies designed to provide support for, although not definitive evidence for, existing hypotheses.
§ Hypothesis-testing studies are studies designed to evaluate in detail hypotheses raised elsewhere.

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 presented, 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 the 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. 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 type of data, as we did in the chapters themselves.

Comparative characteristics of pharmacoepidemiologic data resources

Table 23.1 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. Medicaid databases are the next largest, with the commercial databases approaching that. The UK electronic medical record databases would be next in size, as would the health maintenance organizations, depending on how many are included. The Canadian provincial databases again could be equivalently large, depending on part on how many are included in a study. The other data resources are generally smaller. Case–control surveillance, as conducted by the Slone Epidemiology Unit, can cover a variable population, depending on the number of hospitals and metropolitan areas they include in their 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 find 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.

As regards *relative speed* to completion of the study, 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. US Medicare, Prescription–Event Monitoring in the UK, the provincial health databases in Canada, and the pharmacy-based medical record linkage systems each 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. Randomized trials include populations limited by the various selection criteria plus their willingness to volunteer for the study. The GPRD and THIN use a non-random, large subset of the total

UK population, and so may be representative. Health maintenance organizations (HMOs) 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 common elderly. Some of the remaining data collection approaches or resources are characterized in Table 23.1 as "variable," meaning their representativeness depends on which hospitals are recruited into the study. *Ad hoc* studies are listed in Table 23.1 "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 in 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 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 are not population-based; they include only the specific hospitals that belong to the system. Thus, a patient diagnosed with and treated for venous thromboembolism in a participating hospital could be readmitted to a different, non-participating, 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 which use data from organized health-care 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 21).

Whether cohort studies are possible within a particular data resource would depend on 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 on 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 health-care 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 on reimbursement. These data are likely to be accurate, therefore, 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 42). In addition, there are drugs that may fall beneath a patient's deductibles or co-payments, or not be on formularies. Also, since drug benefits vary depending on the plan, pharmacy files may not capture all prescribed drugs if beneficiaries reach the drug benefit limit. In the UK medical record systems, drugs prescribed by physicians other than the general practitioner could be missed, although continuing 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 on the condition studied, type of drug taken, the questioning technique used, etc.[8–16] (see Chapter 41).

The *validity of the outcome data* is also most certain in hospital-based settings, in which the patient is subjected to intensive medical surveillance. 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 and procedures as markers of the disease and obtaining primary medical records. The outcome data from automated databases are listed as variable, therefore, depending on 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 are probably more accurate.

Control of confounding refers to the ability to control for confounding variables. The most powerful approach to controlling for confounding is randomization. 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 the next most effective. These include HMOs, 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 for control of confounding.

Relatively few of the data systems have data on *inpatient drug use*. The exceptions include spontaneous reporting systems, the in-hospital databases, 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, and so they are less likely to be complete; although the diagnosis can often be validated using medical records, it generally needs to be identified and reported 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 23.1.

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 23.2).

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 HMOs/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, being built primarily for hypothesis generation (see Chapter 30). In the future, new approaches using the internet (e.g., health websites with consumer posting boards) could potentially be used for hypothesis generation of events, including those not coming to medical attention.

Hypothesis-strengthening studies are studies 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. Hypothesis-strengthening studies 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. Alternatively, hypothesis-strengthening studies can be more detailed studies, 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 re-analyzed, 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 the 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 studies 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 exception—see Wei *et al.*[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. (On the other hand, studies of dose–response, duration–response, drug–drug interactions, determinants of response, etc. are more readily done in non-randomized than randomized studies.) Techniques that 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 that allow access to primary records but not the patient are the 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 effects. Studies of *unexpected beneficial effects* are exactly analogous to studies of unexpected adverse effects, in terms of their implications to one's choice of an 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 that usually is related to the outcome under investigation in the study. This issue is discussed in detail in Chapter 37. As described there, it *is* sometimes possible to address these questions using non-experimental 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 the 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, HMOs/ health plans, or Medicaid databases, which cover a large proportion of the United States, 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 36), but this can be very 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 HMOs, or Medicaid databases, which cover a large proportion of the United States, 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 that 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 that 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 HMOs. 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 can be specifically designed to be certain this information is collected. *Ad hoc* studies can be designed to try to collect such information from patients. Finally, occasionally one could collect information on such an outcome in a spontaneous reporting system, if the report came from a patient or if the report came from a health-care provider who became aware of the problem while the patient was visiting for medical care for some other problem. In the future, as noted above, 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 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 HMOs have even longer follow-up time available. However, as HMOs 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 the 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 unlikely, 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 for this, as new drugs are automatically and immediately covered, and in fact reports are much more common in the first 3 years after a drug is marketed. The major databases are the next most useful, especially the commercial databases, 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 especially Medicaid) have a delay in processing 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 that database; of course, if the drug in question is not on the formulary in question, it cannot be studied. 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 were 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 non-steroidal 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., Medicaid, Canadian provincial databases). Because of confounding by cigarette smoking, alcohol, etc., 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, etc.

Alternatively, one might want to address a similar 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, 5 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, HMOs, 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, 3 years is a cause of an extremely rare but serious illness, such as aplastic anemia. One's considerations would be similar to those above, 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. Intentsive 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 (accessed June 2011).
4. Maddison P, Kiely P, Kirkham B, Lawson T, Moots R, Proudfoot D, *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. Anon. 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, Tockman MS, Tonascia S, Rutledge A, *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 Special Applications of Pharmacoepidemiology

CHAPTER 24
Studies of Drug Utilization

David Lee[1] *and Ulf Bergman*[2]

[1] Center for Pharmaceutical Management, Management Sciences for Health, Arlington, VA, USA

[2] Division of Clinical Pharmacology, Department of Laboratory Medicine, Karolinska Institute, WHO Collaborating Centre for Drug Utilization Research and Clinical Pharmacological Services, and Centre for Pharmacoepidemiology, Karolinska University Hospital-Huddinge, Stockholm, Sweden

Introduction

Definitions

Drug utilization is defined by the World Health Organization (WHO) as the "marketing, distribution, prescription and use of drugs in a society, with special emphasis on the resulting medical, social, and economic consequences."[1] Some authors have suggested that the development of drugs relative to health priorities should also be included.[2] This broad definition differs from the more narrow one which appeared in the North American literature, "the prescribing, dispensing and ingesting of drugs."[3,4]

In both of the above definitions, recognition is granted, explicitly or implicitly, of the non-pharmacologic (e.g., socioanthropological, 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 definition of the WHO goes beyond the "process" or "pharmacokinetic" aspect of drug utilization, which is the movement of drugs along the therapeutic drug chain, to include consideration of the various "outcomes" or "pharmacodynamics" of drug use.[5] According to this definition, studies of drug utilization include not only studies of the medical and non-medical factors influencing drug utilization, but also the effects of drug utilization at all levels. 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.[6]

As stated by Lunde and Baksaas,[7] 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."

This chapter focuses on the current status of descriptive epidemiologic approaches to the study of the processes (or "pharmacokinetics") of drug utilization. The epidemiologic approaches to the study of the effects (or "pharmacodynamics") of drug utilization, both beneficial and harmful, are covered elsewhere in this book.

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

Types of drug utilization studies and their uses

Drug utilization studies may be *quantitative* or *qualitative*. In the former, the objective of the study is to quantify the present state, the developmental trends, and the time course of drug usage at various levels of the health-care 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 under-utilization. They also can be used as denominator data for calculating rates of reported adverse drug reactions; 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); to monitor the effects of informational and regulatory activities (e.g., adverse events alerts, delisting of drugs from therapeutic formularies); 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.[2]

Qualitative studies, on the other hand, assess the appropriateness of drug utilization, usually by linking prescription data to the reasons for the drug prescribing (see also Chapter 25). The crucial difference between qualitative and quantitative drug utilization studies is that qualitative drug utilization studies include the concept of appropriateness.[8] 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 upon 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.[9] 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, ecological, and economic consequences.[10–18]

In Spain, the appropriateness of drug utilization has been assessed on the basis of adequate evidence for the clinical efficacy ("high intrinsic value") of the most commonly sold drugs. The analysis revealed a striking proportion of drugs of "doubtful, no, or unacceptable value," among the 400 top pharmaceutical products in sales, albeit a trend toward more apparently rational consumption as reflected in consumption of drugs of "high intrinsic value".[19] This approach has been used to assess: prescribing patterns in France, Germany, Great Britain, Italy,[20] and Spanish primary care centers;[21,22] appropriateness of non-prescription drug sales in Brazil;[23] and the top 50 products sold in Peru.[24]

Another approach analyzed the number of drugs that accounted for 90% of drug utilization (DU90%) and the percentage of these drugs that were consistent with the evidence-based guideline issued by the Drug Committee in the catchment area.[25,26] 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 health-care centers in Stockholm; adherence to the guideline varied between 56% and 74%. The Swedish Medical Quality Council has recommended the DU90% method for assessing quality in drug prescribing. Using the DU90% method, researchers in the Netherlands did not find any 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, duration of treatment (e.g. antidepressants) may be more relevant for quality prescribing than using the drug of first choice in the guidelines; for diabetes, co-medication with HMGCoA reductase inhibitors 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.[26] DU90% has also been used to compare non-

Table 24.1 Drug utilization studies in perspective: operational concepts

	Drug statistics	Drug utilization study	Drug utilization review program
Synonyms (therapeutic)	Drug utilization data	Drug utilization review or drug utilization review study	Drug audit
Quantitative approach	Yes	Usually	Usually
Qualitative approach	No	Maybe	Yes
Continuous (ongoing)	Usually	No	Yes

steroidal anti-inflammatory drug prescribing in Denmark, Italy, Croatia, and Sweden,[27,28] and antibiotics in Denmark and Italy,[29] general intensive care unit antibiotic prescribing and cost patterns in Israel,[30] and to assess the effect of financial incentives linked to self-assessment of prescribing patterns in Swedish primary care.[31]

DUR studies are activities aimed at detection and quantification of problems. They should be distinguished from DUR *programs* (Table 24.1). DUR studies are usually one-time projects, not routinely conducted. They provide for 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 health-care institution. The quality of drug prescribing is evaluated by employing predetermined standards for initiating administrative or educational interventions to modify patterns of drug use which are not consistent with these standards. The measurement of the effectiveness of these interventions is an integral part of the program.[8,32]

In the US, DUR programs (commonly known in hospitals as Drug Use Evaluation or DUE Programs) are part of the quality assurance activities required by Medicaid–Medicare regulations, the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), the former Professional Standards Review Organizations (PSRO), and Section 4401 of the Omnibus Budget Reconciliation Act of 1990.[32] 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.[33–35] 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).

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, drugs often fail to live up to their potential, with consequent morbidity and mortality. Even

when used appropriately, drugs have the potential to cause harm. However, a large proportion of their adverse effects is predictable and preventable.[36]

Adverse drug reactions and drug non-adherence are important causes of adult and pediatric hospital admissions[37,38] (see also Chapter 45). Many of these drug-related admissions may be preventable through the application of existing principles and data.[39] 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 adverse drug interactions; 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 non-prescription 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 the prescribed drugs.[36] Increased morbidity or mortality due to medication error[40] (see Chapter 45), poor patient adherence[41] (see Chapter 42), discontinuation of therapy,[42–44] and problems in communication resulting from modern day fragmentation of patient care are also to be considered (see Chapter 25). 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 that in other areas.[45] In the US, the underuse of beta-blockers in elderly patients with myocardial infarction was associated with an increased risk of death.[46] Other studies have documented significant underuse of antithrombotic drugs,[47–49] lipid-lowering therapy,[44,50,51] beta-blockers,[52] aspirin,[53] and thrombolytics,[54] in patients with appropriate indications, but outcomes were not assessed.

Therapeutic practice, as recommended by relevant professional bodies, academic researchers, and opinion leaders is initially based predominantly on data from premarketing clinical trials. Complementary data from clinical experience and studies in the postmarketing period may result in changes in indication (e.g., antibiotic no longer 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 non-steroidal antinflammatory agents) among others.[55,56] 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.[57]

Methodologic problems to be addressed by pharmacoepidemiologic research

A considerable amount of drug use data may be obtainable or are already available, the usefulness of which depends on the question at hand. All have certain limitations in their direct clinical relevance.[58] 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, the diagnosis or indication, and the dose. The data available are only approximations of this for reasons that are described below, and thereby raise many questions about their presentation and interpretation. For qualitative studies, the ideal is a count of the number of patients in a defined population who use a drug inappropriately during a particular time frame, of all those who received the drug in that population during that time frame. However, such drug exposure and diagnosis data may not be routinely captured or reported. In addition, the criteria to be used to define "appropriate" are arbitrary.

Since most statistics on drug consumption were compiled for administrative or commercial reasons,

Table 24.2 Types of drug utilization data available

1	Cost or unit cost
2	Weight
3	Number of tablets, capsules, doses, etc.
4	Number of prescriptions
5	Number of patients ingesting drug*

*Generally not available.

the data are usually expressed in terms of cost or volume (Table 24.2). 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 the amount of drug exposure in the population. Moreover, cost data are influenced by price fluctuations over time, distribution channels, inflation, exchange rate fluctuations, price control measures, etc.[59]

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 sizes also vary, so it is difficult to translate the number of tablets into the number of exposed patients.

The number of prescriptions is the measure most frequently used in drug utilization studies. However, different patients receive a different number of prescriptions in any given time interval. 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 all these units 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 profiles of physicians according to the number, monetary value, and even type of prescription ordered during a given time period. Pharmacies may 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 very 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. Likewise, cost profiles are not necessarily indicative of appropriateness, whether high or low relative to the mean.

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 physiological status, sex, habits), drug dosage (over- or under-dosage), concomitant diseases (that might contraindicate or interfere with chosen therapy), and the use of other drugs (interactions). However, no single source is generally available for obtaining all this information. Moreover, because of incompleteness, the medical record may not be a very useful source of drug use data.[60,61]

Generally agreed upon standards or criteria for appropriateness, based upon currently available knowledge, are essential elements of the drug

utilization review process. These criteria must be: based on scientifically established evidence; updated regularly according to new scientific evidence; explicitly stated (to ensure consistency in the evaluations); and applicable to a given setting.[62] The development and standardization of these criteria are major undertakings. Finally, for drug utilization review programs, even the strategy to be used to optimize one's intervention is still unclear.

Currently available solutions

The evolution of drug utilization studies

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 the delayed adverse effects, and increasing concern about the cost of drugs, as reflected in the increase in both the monetary sales and the volume of drug prescriptions.[63,64] However, the development of pharmacoepidemiologic methods can be characterized by two different lines of work (drug utilization studies as performed in Europe versus as performed in the US), approaching each other from opposite directions, strongly influenced by the varied availability and accessibility of data sources.

Drug utilization studies at the national and international levels have been more developed in Europe, where this line of research was pioneered by the Scandinavian countries, Scotland, and Northern Ireland. Under the auspices of the WHO Regional Office for Europe, a Drug Utilization Research Group was established in the 1970s to stimulate interest in comparative studies with a common methodology.[63] 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 to 3000 in Norway and Sweden), and the availability of centralized statistics on sales or dispensed prescriptions.[63] Drug utilization studies in Europe have been predominantly quantitative, describing and comparing patterns of utilization of specific groups of drugs according to geographic regions and time. For example, international studies have documented wide variations in the utilization of antidiabetic,[63,65] psychotropic,[33] NSAIDs,[27,28] antihypertensive,[33,63] antibiotic,[66] and lipid-lowering drugs[67] among European and other countries. Follow-up studies on the utilization of antidiabetic and antihypertensive drugs among some of these countries indicate that the differences cannot be explained only by differences in the prevalence of disease.[68–71] National studies have also revealed striking variations in drug utilization among regions and communities within the same country.[33,63] One study addressed the relationship between variations in drug sales and treatment outcomes. In particular, the degree of good metabolic control, based on body mass index and glycosylated hemoglobin, in diabetic subjects was assessed in three Swedish areas with high, medium, and low sales of antidiabetic drugs. Regardless of sales volume, good metabolic control was found among only 16%, 17%, and 12% of subjects, respectively.[72]

In Canada and the US, drug utilization research has developed on a smaller scale, primarily at institutional or local health program levels. Factors that have hindered studies at a national level have been the size of the population, the number of pharmaceutical products on the market (2000 to 30 000), and the lack of an all-encompassing pharmaceutical data collection system.[73] Data on drug use are more readily available from health plans, health delivery institutions, and public health-care 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.[74] 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.[10–18] This was followed by studies that have targeted medications for cardiovascular diseases.[46–54] While only one study described the national pattern of drug utilization

and expenditures in the US in 1982,[73] over the past 10 years several studies have addressed the use of various types of medications, including herbal and other natural products, in adults and children.[75–78]

Because of the critical importance of decision making in drug prescribing, a number of studies have addressed the factors that influence this decision: education, advertising, colleagues, working circumstances, personality, control and regulatory measures, and demands from society and patients.[79,80] 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 excessive or 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.[81]

The intervention strategies aimed at improving prescribing behavior in hospital as well as primary care settings have been critically reviewed.[81–85] These may include (discussed in Chapter 25): 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.

Current data sources

Currently available computer databases for studies of drug utilization may be classified as non-diagnosis-linked and diagnosis-linked (Table 24.3). Most of these data sources lack information on morbidity and are mostly used for generating drug utilization statistics and descriptive studies of patterns of drug consumption. Some collect data in the form of drug sales (e.g., the Danish Medicines Agency, the Norwegian Institute of Public Health, and the National Corporation of Pharmacies in Sweden, published regularly on the respective web sites: www.laegemiddelstyrelsen.dk, www.legemiddelforbruk.no, www.apoteket.se); drug movement at various levels of the drug distribution channel (IMS America's National Prescription Audit, US Pharmaceutical Market-Hospitals, US Pharmaceutical Market-Drugstores; www.imshealth.com); pharmaceutical or medical billing data (Prescription Pricing Authority in the UK, Spain's Drug Data Bank, Medicaid Management Information System);[63,64,86,87] or all prescriptions dispensed (National Corporation of Pharmacies in Sweden, www.apoteket.se).

The County of Jämtland Project (Sweden) is of interest for longitudinal patient-specific studies of drug utilization.[45,88,89] All drug prescriptions dispensed to 17 000 patients (14% of Jämtland's population) have been continuously monitored since 1970. The recorded information includes: the patient's unique identity number; name, dosage, quantity, and price of the drug; date of dispensing; dispensing pharmacy; and prescribing physician. Information relating to morbidity (diagnoses), however, is missing. Similar individual-linked drug use data is also available from many local health systems covering populations of 300 000 to 500 000 inhabitants in Italy; these databases may provide data on incidence and prevalence of drug use.[29]

The Odense Pharmacoepidemiologic Database (OPED) and the Pharmacoepidemiologic Prescription Database of the County of North Jutland are two similar databases which include about half a million inhabitants in Denmark.[90] These databases contain all dispensed prescriptions since the early 1990s. The following information is captured for each prescription: a unique person identifier, the date of dispensing, identification of the dispensed product, the pharmacy, and the prescriber. The databases do not include information on over-the-counter medications (e.g., laxatives, analgesics, ibuprofen, antihistamines, antitussives, and certain antiulcer drugs) and non-subsidized drugs (e.g.,

Table 24.3 Some computer databases for drug utilization studies

Not diagnosis-linked	Diagnosis-linked
North America	
National Prescription Audit*	National Disease and Therapeutic Index*
US Pharmaceutical Market—Drugstores*	Kaiser Permanente Medical Plan†
US Pharmaceutical Market—Hospitals*	Group Health Cooperative†
Medicaid Management Information Systems	
	The Slone Survey‡
Saskatchewan Health Plan†	
Europe	
Swedish National Corporation of Pharmacies	Sweden's Community of Tierp Project
Sweden's County of Jämtland Project	United Kingdom's General Practice Research Database
Norwegian Institute of Public Health	The Netherlands' Integrated Primary Care Information Database
United Kingdom's Prescription Pricing Authority	PHARMO Record Linkage System
Spain's Drug Data Bank (National Institute of Health)	
Denmark's Odense Pharmacoepidemiologic Database	
Denmark's County of North Jutland Pharmacoepidemiologic Prescription Database	
Danish Registry of Medicinal Product Statistics	
Finnish Prescription Registry	
Norwegian Prescription Database	
Swedish Prescribed Drug Register	
Icelandic Pharmaceutical Database	

*IMS America, Ltd.
†Patient-specific data available for longitudinal studies.
‡Reason for use.

oral contraceptives, hypnotics, and sedatives). They have been used for a number of population-based pharmacoepidemiologic surveys such as the use of the new antidepressants,[91] inappropriate use of inhaled steroids in asthma treatment,[92] inappropriate use of sumatriptan,[93] hemorrhagic complication during oral anticoagulant therapy,[94] and low use of long-term hormone replacement therapy.[95] The OPED database has also been used to develop a graphical approach to reduce the overwhelming volume of data in population-based pharmacoepidemiologic databases into a few parameters and a "waiting time distribution" which can be used to screen for certain unusual or unexpected patterns of drug use.[44,96] Based on prescriptions dispensed to individual patients, key parameters such as incidence, 1-year and point prevalence, duration of treatment, relapse rate, and seasonality of medication use are analyzed and presented in a graphic format.

In the US, several databases that contain both drug and morbidity data have been used to a relatively limited extent for this type of study, as opposed to studies of drug effect. These include data from the Group Health Cooperative of Puget Sound and the Kaiser Permanente Medical Care Programs, described in more detail in Chapter 12. The Tayside Medicines Monitoring Unit (MEMO), and the General Practice Research Database (GPRD) in the United Kingdom (see Chapter 15) are data-

bases that have been developed primarily for drug safety studies, but have also been used to study drug utilization.[97,98]

The National Disease and Therapeutic Index (NDTI), by IMS America, Ltd., is an ongoing study of physician prescribing which is conducted mainly for use by pharmaceutical companies for marketing.[99] This study employs a rotating sample of office-based physicians who record all patient encounters and corresponding "drug mentions" for 2-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 pharmacological action.[73] Data have been made available to academic researchers (for a fee) and the US Food and Drug Administration.[73] Although useful for studies of prescribing, longitudinal patient-specific studies are not possible with this database.

The Community of Tierp Project is run by the Center for Primary Care Research, University of Uppsala, Sweden. Prescription and morbidity data are routinely collected from all pharmacies and the health center within the community for all residents since 1972.[100] The database has been used to study the use of benzodiazepines, antidepressant drugs,[100] antidiabetic medications,[101,102] and benzodiazepines.[103] It has also been used to study the impact of over-the-counter nasal sprays on sales, prescribing, and physician visits.[104] Limitations of this database are the size of the population covered (21 000 persons) and questions regarding the representativeness of this community for the whole of Sweden.

The Swedish Prescribed Drug Register, established in 2005, contains data with unique patient identifiers for all dispensed prescriptions in ambulatory care.[105] This registry includes data on the patient (age, sex, personal identification number, place of residence), dispensed drug (Anatomic Therapeutic Chemical classification code, Defined Daily Dose number, prescribed dose, package, reimbursement, date of prescribing and dispensing), prescriber (profession, specialty, workplace), and pharmacy (identifier, location). Four other Nordic databases (the Danish Registry of Medicinal Product Statistics, the Finnish Prescription Registry, the Icelandic Pharmaceutical Database, the Norwegian Prescription Database) contain similar data and have potential for studies that link two or more databases.[106]

The Integrated Primary Care Information (IPCI) Database, established at Erasmus University in the Netherlands, 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 the 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. This database has been used to study the use of preventive strategies in patients receiving non-steroidal anti-inflammatory agents,[107] and trends in primary care prescribing for heart failure.[108] The PHARMO system, another database that links community pharmacy and hospital data, is discussed in Chapter 18.

In Canada, the province of Saskatchewan has a series of computerized databases describing health services paid for by the provincial Department of Health, including prescription drugs.[109] A variety of drug utilization studies have been performed using these data, which are described in more detail in Chapter 17.

In the US, Medicaid medical and pharmaceutical billing data have been available for drug utilization studies. The most frequently used databases for academic pharmacoepidemiologic research are discussed in Chapter 14. The Protocare Sciences Proprietary Medicaid Database (formerly COMPASS[R]) and DURbase[R], both originally developed by Health Information Designs, Inc.), are examples of databases that are used for drug utilization review programs serviced by commercial firms. Drug utilization studies performed using COMPASS[R] have been limited.[110,111] Medicaid data are now frequently obtained from sources other than these commercial vendors (see Chapter

14). With the disadvantaged and disabled population included in Medicaid, however, the generalizability of the results is a potential concern, especially for such descriptive studies.

The Slone Epidemiology Unit of Boston University has developed a novel population-based database that includes prescription and non-prescription drugs, vitamins/ minerals, and herbal/ supplements.[75] From 1998 through 2007, the Slone Epidemiology Unit conducted a telephone survey of a random sample of non-institutionalized continental US population (48 states and the District of Columbia). The survey excluded individuals without home telephones, those residing temporarily in vacation homes, nursing homes, rehabilitation hospitals, and individuals in prisons, military barracks, or college/ university dormitories without telephones in individual rooms. Information was collected on each medication used at any time during the 7 days preceding the phone interview, the reason for use, number of days that medication was taken, and total duration of use. Information on dose and number of pills taken is collected for medications containing acetylsalicylic acid, acetaminophen, ibuprofen, or conjugated estrogens. Other information elicited included age, sex, race, Hispanic origin, years of education, income, health insurance prescription coverage, ZIP code of residence, and for women between 18 and 50 years of age, the pregnancy status, including due date or last menstrual period. Data from over 3000 interviews in the first 3 years of the survey suggested that more than 80% of the US adult population took one prescription or non-prescription medication; 25% took multiple products; and 40% took vitamins/ minerals, while 16% took herbals/ supplements.[75] The overall participation rate of 72% of eligible subjects decreased significantly to 51% in 2006; while this is still at the upper range of participation rates achieved by random digit dialing methods, selection bias may be a concern.[112]

Although the use of health insurance databases has also been reported in countries outside North America and Europe,[113–115] medical and pharmaceutical databases are generally not available in most developing 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 early 1990s by the International Network for Rational Use of Drugs (INRUD) and WHO.[116] 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 seven complementary indicators to study drug use in health facilities (Table 24.4). These indicators can be used to describe prescribing

Table 24.4 WHO/INRUD 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 labeled
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

practice,[117] conduct monitoring and supervision,[118] and assess the impact of interventions.[119–121] WHO has compiled indicator results and other findings reported in studies conducted in 97 developing and transitional countries between 1990 and 2006.[122]

INRUD has recently 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.[123,124] The four indicators were percentage of patients with self-reported full adherence, percentage of days covered by ARVs dispensed, percentage of records with 30-day gap in ARVs dispensed, and percentage of patients who attended within 3 days of scheduled appointment. These indicators allow assessment and comparison of programs and facilities, and monitoring and evaluation of interventions.

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.[125] The DDD is the assumed average daily maintenance dose for a drug for its main indication in adults. Expressed as DDDs per 1000 inhabitants per day, for chronically used drugs, it can be interpreted as the proportion of the population that may receive treatment with a particular medicine on any given day. For use in hospital settings, the unit is expressed as DDDs per 100 bed-days (adjusted for occupancy rate); it suggests the proportion of inpatients that may receive a DDD. For medicines that are used 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. The method has been useful in describing and comparing patterns of drug utilization,[35,63,64] providing denominator data to estimate reported adverse drug reaction rates,[126] performing epidemiologic screening for problems in drug utilization,[35] and monitoring the effects of informational and regulatory activities.[127,128] It has also been used to study variations in antimicrobial utilization,[66,129] antimicrobial utilization and their correlation with antimicrobial resistance in outpatient[130] and inpatient settings in Europe,[131] and report on sustained reduction of antibiotic use and low bacterial resistance with implementation of a multidisciplinary, coordinated national antimicrobial and rational use program.[132]

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 health-care 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 Scandinavia and is increasingly used by researchers in other regions.[127,133–140] A WHO manual on drug utilization research provides an overview of the method.[141] Concerned with increasing antimicrobial resistance, the European Surveillance of Antimicrobial Consumption (ESAC) has recently proposed 12 DDD-based drug-specific quality indicators for outpatient antibiotic use.[142] These indicators were proposed as signals that trigger further detailed analysis of potential problems.

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 DDD 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,[143] the concept and its applicability have not been incorporated into the WHO method.[141] 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 DDDs. The PDD is the average daily dose prescribed, as obtained from a representative sample of prescriptions.[144] 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.[145] For certain groups of drugs, such as the oral antidiabetics, the mean PDD may be lower than the corresponding DDDs. Up to two-fold variations in the mean PDD have been documented in international comparisons.[144] Higher PDDs have been observed in the US relative to Sweden for commonly prescribed drugs, such as hydrochlorothiazide, diazepam, and oxazepam.[146–148] In risk assessments of antidepressants among suicides, a refined person-year of use estimate was obtained from adjusting the DDD by the average PDD for individual antidepressants.[149] 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.

Classification systems

The Anatomic Therapeutic Chemical (ATC) classification system is generally used in conjunction with the DDD method.[125,141] 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 24.5. 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.

Table 24.5 ATC and 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)

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 glau-

coma) 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 and 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.[150]

The European Drug Utilization Research Group (EuroDURG), formerly WHO Drug Utilization Research Group and currently an association of European national Drug Utilization Research Groups, recommends the use of the ATC classification system for reporting drug consumption statistics and conducting comparative drug utilization research. Australia (www.health.gov.au), Denmark (www.laegemiddelstyrelsen.dk), Finland (www.kela.fi), Iceland (see www.statice.is), Norway (www.legemiddelforbruk.no), and Sweden (www.apoteket.se) produce annual reports on drug consumption and make them available in print and/or web-based electronic versions. The WHO International Drug Monitoring Program uses the system for drug coding in adverse drug reaction monitoring (www.who-umc.org). Some developing countries have begun to use the ATC system to classify their essential drugs;[151,152] this may eventually lead to preparation of annual drug utilization statistics.[153]

In the US, the Iowa Drug Information System (IDIS) is a hierarchical drug coding system that is based on the three therapeutic categories of American Hospital Formulary Society (AHFS), to which a fourth level was added to code individual drug ingredients.[154] The IDIS code has eight numeric digits, two digits per level (see Table 24.5). This coding system was used in the Established Populations for Epidemiologic Studies of the Elderly survey.[154] Other coding systems such as the National Drug Code and the Veterans' Administration Classification[155] do not provide unique codes for drug ingredients.

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 qualitative drug utilization studies, and are discussed more in Chapter 25. Two innovative intervention strategies 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[156] 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.*[157] and Ray *et al.*[158] 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.[111,159,160] 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 that the prescriber was unaware of. 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.[159] However, the interpretation of these results is hampered by the use of a non-experimental design. Other authors have found no effect on measures of prescribing or on patient outcomes.[161] 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.[81]

Conceptually, DUR programs are aimed at the improvement of medical care and cost-containment. However, in practice traditional approaches have focused on the control of 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.*,[111] Morse *et al.*,[159] and Groves[160] 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 health-care system which is much larger than the cost of the drugs themselves. (This is discussed more in Chapter 38.)

Despite their appeal, the effectiveness of DUR programs still remains to be established. A study of six Medicaid programs failed to identify an effect of retrospective drug utilization review on the rate of potential prescribing errors and rate of all-cause or specific cause hospitalizations.[161] 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.[162]

The future

Opportunities

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. Analysis of medicine use by gender and age group may suggest important associations, as seen in a study on antidepressant medication use and decreased suicide rates.[163] The increasing availability of population-based data resources will facilitate studies of incidence and prevalence of medicine use by age and gender, such as those conducted in Sweden and Denmark.

Numerous studies have addressed the factors influencing drug prescribing. However, the relative importance of the many determinants of appropriate prescribing still remains 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.[164] Other approaches in addition to educational and informational measures are being explored.

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.[81,82] Early studies conducted in Australia[165] and Denmark[166] 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 25), 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 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,[156] Schaffner *et al.*,[157] and Ray *et al.*[158]

More clinically applicable approaches to drug utilization review 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[159,160,167,168] 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, health-care policy options should be based on sound scientific evidence.[169]

Problems

The use of computerized databases has greatly facilitated the study 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[170] is still unavailable, although there is increasing use of electronic medical records for routine practice in countries such as the Netherlands, Australia, the United Kingdom, and the US. Existing medical and pharmaceutical databases, with all their described limitations, will continue to be the main resources for these drug utilization studies.

Confidentiality of patient records has been successfully handled at the technical level. However, in many countries political acceptance may be much more difficult to achieve. EuroDURG researchers reported difficulties arising from confidentiality laws in five of 10 European countries.[171] It was feared that implementation of Directive 95/46/EC of 24 October 1995, on the protection of individuals with regard to the processing of personal data and on the free movement of such data in the European Union, may adversely affect researcher access to patient health data (see also Chapter 35). However, these concerns can be addressed through procedures that are consistent with the guidelines for Good Practice in Data Privacy, Medical Record Confidentiality, and Research developed by the International Society of Pharmacoepidemiology (ISPE).[172] For example, in Sweden, personally identified health data may only be used for research, statistics, and epidemiologic studies. Registry data cannot be used to monitor performance of health professionals or for administrative purposes that may influence those being registered. Research using patient data requires ethics committee approval which may additionally ask for research subject formal consent. Anonymous data are provided with requirements to keep the data safe, not share with others without prior approval from the data agency, and discard the data at the end of the study or when they are no longer needed.[105]

In many countries research is usually not awarded high priority, resulting in reduced opportunities for financing much needed drug utilization research. 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),[173] the International Clinical Epidemiology Network (INCLEN) (www.inclentrust.org),[174] the European Drug Utilization Research Group (EuroDURG) (www.eurodurg.com),[175] the Latin American Group for

Drug Epidemiology (DURG-LA),[176] and the International Network for Rational Use of Drugs (INRUD) (www.inrud.org/).[177,178]

In summary, the study of drug utilization continues to evolve. The development of computerized databases that allow the linkage of drug utilization data to diagnoses, albeit subject to some inherent limitations, is contributing to expansion of this field of study. 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[179] and developing countries.[180] Drug utilization review 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 under-unexplored, but the political issue regarding the confidentiality of medical records, as well as limitations in funding and manpower will determine the pace of growth of drug utilization research.

References

1. WHO Expert Committee. *The Selection of Essential Drugs.* Technical Report Series No. 615. Geneva: World Health Organization, 1977.
2. Lunde PKM, Andrew M, Baksaas I. Drug utilization-an instrument in drug research. In: Kewitz H, Roots I, Voight K, eds. *Epidemiological Concepts in Clinical Pharmacology*. Berlin: Springer-Verlag, 1987, pp. 57–72.
3. Brodie DC. *Drug Utilization and Drug Utilization Review and Control*. NCHS-RD-70-8. Rockville: Department of Health, Education, and Welfare, National Center for Health Services Research and Development, Health Services and Mental Health Administration, 1970.
4. Conley BE. *Social and Economic Aspects of Drug Utilization Research*. Hamilton: Drug Intelligence Publications, Inc., 1976.
5. Baksaas I, Lunde PKM. Drug utilization: Pharmacokinetics in the community. *Trends Pharmacol Sci* 1981; **2**: 5–7.
6. Laporte JR, Tognoni G, eds. *Principios de Epidemiología del Medicamento*, 2nd edn. Barcelona: Ediciones Científicas y Técnicas, 1993.
7. Lunde PKM, Baksaas I. Epidemiology of drug utilization—basic concepts and methodology. *Acta Med Scand Suppl* 1988; **721**: 7–11.
8. Stolar MH. Drug use review: Operational definitions. *Am J Hosp Pharm* 1978; **35**: 76–8.
9. Lee PR. America is an overmedicated society. In: Lasagna L, ed. *Controversies in Therapeutics*. Philadelphia: W. B. Saunders Company, 1980, pp. 4–15.
10. Scheckler NE, Bennet JV. Antibiotic usage in seven community hospitals. *JAMA* 1970; **213**: 264–7.
11. Roberts WA, Visconti AJ. The rational and irrational use of systemic antimicrobial drugs. *Am J Hosp Pharm* 1972; **29**: 828–34.
12. Castle M, Wilfert CM, Cate TR, Osterhout S. Antibiotic use at Duke University Medical Center. *JAMA* 1977; **237**: 2819–22.
13. Perry TL, Guyatt GH. Antimicrobial drug use in three Canadian hospitals. *CMAJ* 1977; **116**: 253–6.
14. Achong MR, Hauser BA, Krusky JL. Rational and irrational use of antibiotics in a Canadian teaching hospital. *CMAJ* 1977; **116**: 256–9.
15. Shapiro M, Townsend TR, Rosner B, Kass EH. Use of antimicrobial drugs in general hospitals. II. Analysis of patterns of use. *J Infect Dis* 1979; **139**: 698–706.
16. Townsend TR, Shapiro M, Rosner B, Kass EH. Use of antimicrobial drugs in general hospitals. IV. Infants and children. *Pediatrics* 1979; **64**: 573–8.
17. Schollenberg E, Albritton WL. Antibiotic misuse in a pediatric teaching hospital. *CMAJ* 1980; **122**: 49–52.
18. Naqvi SH, Dunkle LM, Timmerman KJ, Reichley RM, Stanley DL, O'Connor D. Antibiotic usage in a pediatric medical center. *JAMA* 1980; **242**: 1981–4.
19. Laporte JR, Porta M, Capellá D. Drug utilization studies: A tool for determining the effectiveness of drug use. *Br J Clin Pharmacol* 1983; **16**: 301–4.
20. Garattini S, Garattini L. Pharmaceutical prescriptions in four European countries. *Lancet* 1993; **342**: 1191–2.
21. Casado Zuriguel I. Intervención para mejorar la calidad de prescripción de antibióticos en una area básica de salud. *Aten Primaria* 1993; **11**: 37–9.
22. Mata Cases M, Avellana Revuelta E, Davins Miralles J, Calvet Junoy S, Hortelano García MA, Martín López A. Mejora continua de la calidad de la pre-

scripción crónica en un centro de atención primaria: seguimiento de 5 años. *Aten Primaria* 1994; **13**: 172–6.
23. Heineck I, Schenkel EP, Vidal X. Medicamentos de venta libre en el Brasil. *Rev Panam Salud Pública* 1998; **3**: 385–91.
24. Meza Cornejo EA. *Valor Terapéutico de los Medicamentos más Vendidos en ele Perú*. AIS LAC Serie Investigaciones y Monografías No. 31/2010. Lima, Perú: Acción Internacional para la Salud, 2010.
25. Wettermark B, Pehrsson Å, Jinnerot D, Bergman U. Drug utilization 90% profiles—a useful tool for quality assessment of prescribing in primary health care in Stockholm. *Pharmacoepidemiol Drug Saf* 2003; **12**: 499–510.
26. 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**: 1171–77.
27. 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**: 269–72.
28. Vlahovic-Palcevski 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**: 209–14.
29. 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 Chemotherapy* 2002; **50**: 989–97.
30. Krivoy N, El-Ahal WA, Bar-Lavie Y, Haddad S. Antibiotic prescription and cost patterns in a general intensive care unit. *Pharmacy Practice* 2007; **5**: 67–73.
31. Wettermark B, Pehrsson Å, Juhasz-Haverinen M, Veg A, Edlert M, Tornwall-Bergendahl, *et al*. *Quality in Primary Care* 2009; **17**: 179–89.
32. Todd M. Drug use evaluation. In: Brown TR, ed. *Handbook of Institutional Pharmacy Practice*, 3rd edn. Bethesda, MD: American Society of Hospital Pharmacists, 1992, pp. 261–71.
33. Sjöqvist F, Agenäs I, eds. Drug utilization studies: Implications for medical care. *Acta Med Scand* 1984; Suppl 683.
34. Baksaas I, Lunde PKM. National drug policies: The need for drug utilization studies. *Trends Pharmacol Sci* 1986; **7**: 331–4.
35. Westerholm B. Therapeutic auditing at the national and international level. *Br J Clin Pharmacol* 1986; **22**: 55S–59S.
36. Melmon KL. Preventable drug reactions: Causes and cures. *N Engl J Med* 1971; **284**: 1361–3.
37. 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**: 77–83.
38. Runciman WB, Roughead EE, Semple SJ, Adams RJ. Adverse drug events and medication errors in Australia. *Int J Qual Health Care* 2003; **15** (Suppl 1): i49–59.
39. Davidsen F, Haghfelt T, Gram LF, Brøsen K. Adverse drug reactions and drug non-compliance as primary causes of admission to a cardiology department. *Eur J Clin Pharmacol* 1988; **34**: 83–6.
40. Bates DW. Medication errors: how common are they and what can be done to prevent them? *Drug Saf* 1996; **15**: 303–10.
41. Haynes RB, Taylor DW, Sackett DL, eds. *Compliance in Health Care*. Baltimore: Johns Hopkins University Press, 1979.
42. Andrade SE, Walker AM, Gottlieb LK, Hollenberg NK, Testa MA, Saperia GM, *et al*. Discontinuation of antihyperlipidemic drugs-do rates reported in clinical trials reflect rates in primary care settings? *N Engl J Med* 1995; **332**: 1125–31.
43. Avorn J, Monette J, Lacour A, Bohn RL, Monane M, Mogun H, *et al*. Persistence of use of lipid-lowering medications: a cross-national study. *JAMA* 1998; **279**: 1458–62.
44. 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**: 463–71.
45. 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**: 444–8.
46. 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**: 115–21.
47. Whittle J, Wickenheiser L, Venditti LN. Is warfarin underused in the treatment of elderly persons with atrial fibrillation? *Arch Intern Med* 1997: **157**: 441–5.
48. Antani MR, Beyth RJ, Covinsky KE, Anderson PA, Miller DG, Cebul RD, *et al*. Failure to prescribe warfarin to patients with nonrheumatic atrial fibrillation. *J Gen Intern Med* 1996; **11**: 713–20.
49. Albers GW, Yim JM, Belew KM, Bittar N, Hattemer CR, Phillips BG, *et al*. Status of antithrombotic therapy for patients with atrial fibrillation in university hospitals. *Arch Intern Med* 1996; **156**: 2311–6.
50. Mendelson G, Aronow WS. 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**: 1128–31.
51. Aronow S. Underutilization of lipid-lowering drugs in older persons with prior myocardial infarction and a serum low-density lipoprotein cholesterol >125mg/dl. *Am J Cardiol* 1998; **82**: 668–9, A6, A8.
52. McLaughlin TJ, Soumerai SB, Willison DJ, Gurwitz JH, Borbasi C, Guadagnoli E, *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**: 799–805.
53. Krumholtz HM, Radford MJ, Ellerbeck EF, Hennen J, Meehan TP, Petrillo M, *et al*. Aspirin in the treatment of acute myocardial infarction in elderly Medicare beneficiaries. Patterns of use and outcomes. *Circulation* 1995; **92**: 2841–7.
54. O'Connor GT, Quinton HB, Traven ND, Ramunno LD, Dodds TA, Marciniak TA, *et al*. Geographic variation in the treatment of acute myocardial infarction: the Cooperative Cardiovascular Project. *JAMA* 1999; **281**: 627–33.
55. Haider BA, Saeed MA, 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.
56. Strom BL, Melmon KL, Miettinen OS. Postmarketing studies of drug efficacy. *Arch Intern Med* 1985; **145**: 1791–4.
57. Lasagna L. A plea for the naturalistic study of medicines. *Eur J Clin Pharmacol* 1974; **7**: 153–4.
58. Rucker TD. Data, sources, and limitations. *JAMA* 1974; **230**: 888–90.
59. Anonymous. *Consumption of Drugs: Report on a Symposium. Euro 3102*. Copenhagen: WHO Regional Office for Europe, 1970.
60. Bergman U. Pharmacoepidemiology—from description to quality assessment. A Swedish perspective. *Norwegian J Epidemiology* 2001; **11**: 31–6.
61. Strom BL, Carson JL, Halpern AC, Schinnar R, Snyder ES, Stolley PD, *et al.* Using a claims database to investigate drug-induced Stevens-Johnson syndrome. *Stat Med* 1991; **10**: 565–76.
62. Rucker TD. Drug utilization review: Guidelines for program development. In: Alloza JL, ed. *Clinical and Social Pharmacology. Postmarketing Period*. Aulendorf: Cantor, 1985, p. 57.
63. Bergman U, Grímsson A, Wahba AHW, Westerholm B, eds. Studies in drug utilization. European series No. 8. Copenhagen: WHO Regional Office for Europe; 1979. Anonymous. Panel III. Intervention: the use of drug utilization review outcomes. *Clin Pharmacol Ther* 1991; **50**: 633–5.
64. Dukes MNG, ed. *Drug Utilization Studies: Methods and Uses*. European Series, No. 45. Copenhagen: World Health Organization, Regional Office for Europe, 1993.
65. Bergman U, Elmes P, Halse M, Halvorsen T, Hood H, Lunde PK, *et al*. The measurement of drug consumption: drugs for diabetes in Northern Ireland, Norway and Sweden. *Eur J Clin Pharmacol* 1975; **8**: 83–9.
66. Cars O, Mölstad S, Melander A. Variation in antibiotic use in the European Union. *Lancet* 2001; **357**: 1851–3.
67. 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 (Emilio Romagna Region), Norway and Sweden. *Eur J Clin Pharmacol* 1997: **53**: 185–9.
68. WHO Drug Utilization Research Group (DURG). Validation of observed differences in the utilization of antihypertensive and antidiabetic drugs in Northern Ireland, Norway and Sweden. *Eur J Clin Pharmacol* 1985; **29**: 1–8.
69. WHO Drug Utilization Research Group (DURG). Therapeutic traditions in Northern Ireland, Norway and Sweden. I. Diabetes. *Eur J Clin Pharmacol* 1986; **30**: 513–9.
70. WHO Drug Utilization Research Group (DURG). Therapeutic traditions in Northern Ireland, Norway and Sweden. II. Hypertension. *Eur J Clin Pharmacol* 1986; **30**: 521–5.
71. Groop P-H, Klaukka T, Reunanen A, Bergman U, Borch-Johnsen K, Damsgaard E-M, *et al*.

Diabetesläkemedel i Norden. Analys av orsaker till variationen i förbrukningen. [Antidiabetic drugs in the Nordic countries. Reasons for variation in their use.] Social Insurance Publication ML 105. Helsinki: Social Insurance Institution, 1991.

72. 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**: 12–9.
73. Baum C, Kennedy DL, Forbes MB, Jones JK. Drug use and expenditures in 1982. *JAMA* 1985; **253**: 382–6.
74. Stolley PD, Lasagna L. Prescribing patterns of physicians. *J Chronic Dis* 1969; **22**: 395–405.
75. 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**: 337–44.
76. Kelly JP, Kaufman DW, Kelley KE, Rosenberg L, Anderson TE, Mitchell AA. Recent trends in use of herbal and other natural products. *Arch Intern Med* 2005; **165**: 281–6.
77. Vernacchio L, Kelly JP, Kaufman DW, Mitchell AA. Cough and cold medication use by U.S. children, 1999–2006: Results from the Slone Survey. *Pediatrics* 2008; **122**: 323–9.
78. Vernacchio L, Kelly JP, Kaufman DW, Mitchell AA. Pseudoephedrine use among U.S. children 1999–2006: Results from the Slone Survey. *Pediatrics* 2008; **122**: 1299–304.
79. Hemminki E. Review of literature on the factors affecting drug prescribing. *Soc Sci Med* 1975; **9**: 111–15
80. Christensen DB, Bush PJ. Drug prescribing: Patterns, problems and proposals. *Soc Sci Med* 1981; **15**: 343–55.
81. 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.
82. Soumerai SB, McLaughlin TJ, Avorn J. Improving drug prescribing in primary care: a critical analysis of the experimental literature. *Milbank Q* 1989; **67**: 268–317.
83. Anderson GM, Lexchin J. Strategies for improving prescribing practice. *CMAJ* 1996; **154**: 1013–7.
84. Freemantle N, Harvey EL, Wolf F, Grimshaw JM, Grilli R, Bero LA. Printed educational materials: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2000; **2**: CD000172.
85. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; **342**: 1317–22.
86. Gómez LA, Lobato CI. Evolución de las bases de datos de medicamentos del Ministerio de Sanidad y Consumo. *Inf Ter Seg Soc* 1987; **11**: 229.
87. Avorn J, Soumerai SB. Use of computer-based Medicaid drug data to analyze and correct inappropriate medication use. *J Med Syst* 1982; **6**: 377–8.
88. Boëthius G, Wiman F. Recording of drug prescriptions in the county of Jämtland, Sweden. I. Methodological aspects. *Eur J Clin Pharmacol* 1977; **12**: 31–5.
89. Henriksson S, Boëthius G, Håkansson J, Isacsson G. Indications for and outcome of antidepressant medication in a general population: a prescription database and medical record study, in Jämtland county, Sweden, 1995. *Acta Psychiatr Scand* 2003; **108**: 427–31.
90. Hallas J. Conducting pharmacoepidemiologic research in Denmark. *Pharmacoepidemiol Drug Saf* 2001; **10**: 619–23.
91. Rosholm J-U, Gram LF, Isacsson G, Hallas J, Bergman U. Changes in the pattern of antidepressants use upon the introduction of the new antidepressants: a prescription database study. *Eur J Clin Pharmacol* 1997; **52**: 205–9.
92. Gaist D, Hallas J, Hansen N-CG, Gram LF. Are young adults with asthma treated sufficiently with inhaled steroids? A population-based study of prescription data from 1991 and 1994. *Br J Clin Pharmacol* 1996; **41**: 2885–9.
93. Gaist D, Tsiropoulos I, Sindrup SH, Hallas J, Rasmussen BK, Kagstrup J, *et al*. Inappropriate use of sumatriptan: population based register and interview study. *Br Med J* 1998; **316**: 1352–3.
94. Steffensen FH, Kristensen K, Ejlersen E, Dahlerup JF, Sorensen HT. Major hemorrhagic complications during oral anticoagulant therapy in a Danish population-based cohort. *J Intern Med* 1997; **242**: 4997–503.
95. Olesen C, Steffensen FH, Sorensen HT, Nielsen GL, Olsen J, Bergman U. Low use of long-term hormone replacement therapy in Denmark: a 5-year population-based survey. *Br J Clin Pharmacol* 1999; **47**: 323–8.
96. Hallas J, Gaist D, Bjerrum L. The waiting time distribution as a graphical approach to epidemiologic measures of drug utilization. *Epidemiology* 1997; **8**: 666–70.
97. Donnan PT, Steinke DT, Newton RW, Morris AD; DARTS/MEMO Collaboration. Changes in treatment

after the start of oral hypoglycaemic therapy in Type 2 diabetes: a population-based study. *Diabet Med* 2002; **19**: 606–10.

98. Bromley SE, de Vries CS, Farmer RD. Utilisation of hormone replacement therapy in the United Kingdom: a descriptive study using the general practice research database. *BJOG* 2004; **111**: 369–76.
99. Nash D. National drug and therapeutic index. *PT* 2002; **27**: 530.
100. Bingefors K. Computerised data bases on prescription drug use and health care in the community of Tierp, Sweden: experiences and challenges from a study of antidepressant-treated patients. *Norwegian J Epidemiol* 2001; **11**: 23–9.
101. Isacson D, Ståhlhammar J. Prescription drug use among diabetics—a population study. *J Chronic Dis* 1987; **40**: 651–60.
102. Ståhlhammar J, Berne C, Svärdsudd K. Do guidelines matter? A population-based study of diabetes use during 20 years. *Scand J Prim Health Care* 2001; **19**: 163–9.
103. Isacson D, Carsjö K, Bergman U, Blackburn JL. Long-term use, mortality and migration among benzodiazepine users in a Swedish community: an eight year follow-up. *J Clin Epidemiol* 1992; **45**: 429–36.
104. Lundberg L, Isacson D. The impact of over-the-counter availability of nasal sprays on sales, prescribing, and physician visits. *Scand J Prim Health Care* 1999; **17**: 41–5.
105. Wettermark B, Hammar N, Fored CM, Leimanis A, Olausson PO, Bergman U, *et al*. The new Swedish Prescribed Drug Register—opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiology Drug Saf* 2007; **16**: 726–35.
106. 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* 2009; **106**: 86–94.
107. Sturkenboom MCJM, Burke TA, Dieleman JP, Tangelder MJD, Lee F, Goldstein JL. Underutilization of preventive strategies in patients receiving NSAIDs. *Rheumatology* 2003; **42** (Suppl 3): iii23–iii31.
108. Pont LG, Sturkenboom MC, van Gilst WH, Denig P, Haaijer-Ruskamp FM. Trends in prescribing for heart failure in Dutch primary care from 1996 to 2000. *Pharmacoepidemiol Drug Saf* 2003; **12**: 327–34.
109. Blackburn JL. The use of automated databases in North America. In: Crommelin DJA, Midha KK, eds. *Topics in Pharmaceutical Sciences 1991: Proceedings of the 51st International Congress of Pharmaceutical Sciences of F.I.P.*, 1991 Sep 1–6, Washington D.C., USA. Stuttgart: Medpharm Scientific Publishers, 1992, pp. 405–13.
110. Strom BL, Carson JL, Morse ML, LeRoy AA. The Computerized On-line Medicaid Pharmaceutical Analysis and Surveillance System: a new resource for postmarketing drug surveillance. *Clin Pharmacol Ther* 1985; **38**: 359–64.
111. Strom BL, Morse ML. Use of computerized data bases to survey drug utilization in relation to diagnoses. *Acta Med Scand Suppl* 1988; **721**: 13–20.
112. Anonymous. *Patterns of Medication Use in the United States 2006. A Report from the Slone Survey*. Slone Epidemiology Center. Boston, MA: Boston University. http://www.bu.edu/slone/SloneSurvey/AnnualRpt/SloneSurveyWebReport2006.pdf (accessed Aug 2011).
113. Chen TJ, Chou LF, Hwang SJ. Trends in prescribing proton pump inhibitors in Taiwan: 1997–2000. *Int J Clin Pharmacol Ther* 2003; **41**: 207–12.
114. Pelaez-Ballestas I, Hernandez-Garduno A, Arredondo-Garcia JL, Viramontes-Madrid JL, Aguilar-Chiu. Use of antibiotics in upper respiratory infections on patients under 16 years old in private ambulatory medicine. *Salud Publica Mex* 2003; **45**: 159–64.
115. 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**: 663–8.
116. World Health Organization. *How to Investigate Drug Use in Health Facilities: Selected Drug Use Indicators*. WHO/DAP/93.1. Geneva: World Health Organization, 1993.
117. Trap B, Hansen EH, Hogerzeil HV. Prescription habits of dispensing and non-dispensing doctors in Zimbabwe. *Health Policy Plan* 2002; **17**: 288–95.
118. Sunartono. From research to action: the Gunungkidul experience. *Essential Drugs Monitor* 1995; **2**: 21–2.
119. Santoso B, Suryawati S, Prawitasari JE, Ross-Degnan D. Small group intervention vs formal seminar for improving appropriate drug use. *Soc Sci Med* 1996; **42**: 1163–8.
120. Ofori-Adjei D, Arhinful DK. Effect of training on the clinical management of malaria by medical assistants in Ghana. *Soc Sci Med* 1996; **42**: 1169–76.
121. 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**: 1177–83.

122. World Health Organization. *Medicines use in Primary Care in Developing and Transitional Countries: Fact Book Summarizing Results from Studies Reported Between 1990 and 2006.* WHO/EMP/MAR/2009.3. Geneva, Switzerland: World Health Organization, 2009.
123. Ross-Degnan D, Pierre-Jacques M, Zhang F, Tadeg H, Gitau L, Ntaganira J, *et al*. Measuring adherence to antiretroviral treatment in resource-poor settings: the clinical validity of key indicators. *BMC Health Services Research* 2010; **10**: 42.
124. Chalker JC, Andualem T, Gitau LN, Ntaganira J, Obua C, Tadeg H, *et al*. Measuring adherence to antiretroviral treatment in resource-poor settings: the feasibility of collecting routine data for key indicators. *BMC Health Services Research* 2010; **10**: 43.
125. WHO Collaborating Centre for Drug Statistics Methodology. *Guidelines for ATC Classification and DDD Assignment 2010*. Oslo, 2009.
126. Bergman U, Boman G, Wiholm BE. Epidemiology of adverse drug reactions to phenformin and metformin. *Br Med J* 1978; **2**: 464–6.
127. Lee D, Chaves Matamoros A, Mora Duarte J. Changing patterns of antibiotic utilization patterns in Costa Rica. *APUA Newslett* 1991; **9**: 7.
128. Schussel K, Schulz M. Prescribing of COX-2 inhibitors in Germany after safety warnings and market withdrawals. *Pharmazie* 2006; **61**: 878–86.
129. Wirtz VJ, Dreser A, Gonzales R. Trends in antibiotic utilization in eight Latin American countries, 1997–2007. *Rev Panam Salud Publica* 2010; **27**: 219–25.
130. Bronzwaer SLAM, Cars O, Bucholz U, Mölstad S, Goettsch W, Veldhuijzen IK, *et al*. A European study on the relationship between antimicrobial use and antimicrobial resistance. *Emerg Infect Dis* 2002; **8**: 278–82.
131. Gulbinovic J, Myrback K-E, Bytautiene 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**: 383–8.
132. Mölstad S, Ernell M, Hanberger H, Melander E, Norman C, Skoog G, *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**: 125–32.
133. Nappo S, Carlini EA. Preliminary finding: consumption of benzodiazepines in Brazil during the years 1988 and 1989. *Drug Alcohol Depend* 1993; **33**: 11–7.
134. Chen L, Wang Y, Jin Y. [Study on drug use in elderly outpatients in Beijing]. *Zhonghua Liu Xing Bing Xue Za Zhi* 2001; **22**: 414–7.
135. Truter I, Wiseman K, Kotze TT. The defined daily dose as a measure of drug consumption in South Africa. A preliminary study. *S Afr Med J* 1996; **86**: 675–9.
136. Su TP, Chen TJ, Hwang SJ, Chou LF, Fan AP, Chen YC. Utilization of psychotropic drugs in Taiwan: an overview of outpatient sector in 2000. *Zhonghua Yi Xue Za Zhi (Taipei)* 2002; **65**: 378–91.
137. Vlahovic-Palevski 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**: 287–93.
138. Ansari F. Use of systemic anti-infective agents in Iran during 1997–1998. *Eur J Clin Pharmacol* 2001; **57**: 547–51.
139. 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**: 55–61.
140. Jirón M, Machado M, Ruiz I. Consumo de antidepresivos en Chile entre 1992 y 2004. *Rev Med Chile* 2008; **136**: 1147–54.
141. WHO International Working Group for Drug Statistics Methodology, WHO Collaborating Centre for Drug Statistics Methodology, WHO Collaborating Centre for Drug Utilization Research and Clinical Pharmacology. *Introduction to Drug Utilization Research*. Geneva: World Health Organization, 2003.
142. Coenen S, Ferech M, Haaijer-Ruskamp FM, Butler CC, Stichele RHV, Verheij TJM, *et al*. European Surveillance of Antimicrobial Consumption: quality indicators for outpatient antibiotic use in Europe. *Qual Saf Health Care* 2007; **16**: 440–5.
143. Stanulovic M, Milosev M, Jakovlojevic V, Roncevic N. Epidemiological evaluation of anti-infective drug prescribing for children in outpatient practice. *Dev Pharmacol Ther* 1987; **10**: 278–91.
144. Bergman U, Sjöqvist F. Measurement of drug utilization in Sweden: methodological and clinical implications. *Acta Med Scand* 1984; **105** (Suppl 683): 15–22.
145. Bergman U. Utilization of antidiabetic drugs in the island of Gotland, Sweden: Agreement between wholesale figures and prescription data. *Eur J Clin Pharmacol* 1978; **14**: 213–20.
146. Bergman U, Dahlström M. Användningen av blodtryckssänkande läkemedel i Sverige. (Use of antihypertensive agents in Sweden) In: Berglund G, ed.

Hypertoni 87. Mölndal: Lindgren & Söner, 1988, pp. 22–32.

147. Bergman U, Dahlström M. Benzodiazepine utilization in Sweden and other Nordic countries. In: Lader MH, Davies HC, eds. *Drug Treatment of Neurotic Disorders*. Edinburgh: Churchill Livingstone, 1986, pp. 43–52.
148. Juergens JP, Bergman U, Baum C, Kennedy DL, Dahlstrom M. Use of benzodiazepines in the USA and Sweden. *Drug Information Association Annual Meeting*, Washington, D.C., June 4, 1986.
149. Isacsson G, Holmberg P, Druid H, Bergman U. The utilization of antidepressants—a key issue in the prevention of suicide. An analysis of 5281 suicides in Sweden 1992–1994. *Acta Psychiatr Scand* 1997; **96**: 94–100.
150. Rönning M, Salvesen Blix H, Harbo BT, Strom H. Different versions of the anatomical therapeutic chemical classification system and the defined daily dose—are drug utilization data comparable? *Eur J Clin Pharmacol* 2000; **56**: 723–7.
151. Departamento de Farmacoterapia. *Lista Oficial de Medicamentos*. Costa Rica: Caja Costarricense de Seguro Social, 2002.
152. The National Essential Drugs Committee. *Standard Treatment Guidelines and Essential Drugs List for Primary Health Care*, 2nd edn. Pretoria, South Africa: National Department of Health, 1998.
153. Sameerah SAR, Lian LM, eds. *Malaysian Statistics on Medicines 2006*. Pharmaceutical Services Division and the Clinical Research Centre, Ministry of Health Malaysia, 2006.
154. 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**: 405–11.
155. Anonymous. Appendix IV: VA medication classification system. In: *USP Dispensing Information*, Vol. I: Drug information for the health care professional. Englewood, CO: Micromedex, Inc., 1999, pp. 3045–62.
156. 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.
157. Schaffner W, Ray WA, Federspiel CF, Miller WO. Improving antibiotic prescribing in office practice. A controlled trial of three educational methods. *JAMA* 1983; **250**: 1728–32.
158. Ray WA, Schaffner W, Federspiel CF. Persistence of improvement in antibiotic prescribing in office practice. *JAMA* 1985; **253**: 1774–6.
159. 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.
160. Groves R. Therapeutic drug-use review for the Florida Medicaid program. *Am J Hosp Pharm* 1985; **42**: 316–9.
161. Hennessy SM, Bilker WB, Zhou L, Weber AL, Brensinger C, Wang Y, Strom BL. Retrospective drug utilization review, prescribing errors, and clinical outcomes. *JAMA* 2003; **290**: 1494–9.
162. Kidder D, Bae J. Evaluation results from prospective drug utilization review: Medicaid demonstrations. *Health Care Financ Rev* 1999; **20**: 107–18.
163. Hall WD, Mant A, Mitchell PB, Rendle VA, Hickie IB, McManus P. Association between antidepressant prescribing and suicide in Australia, 1991–2000: trend analysis. *BMJ* 2003; **326**: 1008–12.
164. Soumerai SB, Avorn J, Ross-Degnan D, Gortmaker S. Payment restrictions for prescription drugs under Medicaid. *N Engl J Med* 1987; **317**: 550–6.
165. O'Connell DL, Henry D, Tomlins. Randomised controlled trial of feedback on general practitioners prescribing in Australia. *BMJ* 1999; **318**: 507–11.
166. Söndergaard J, Andersen M, Stövring H, Kragstrup J. Mailed prescriber feedback in addition to a clinical guideline has no impact: a randomized, controlled trial. *Scand J Prim Health Care* 2003; **21**: 47–51.
167. Monane M, Mathias 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**: 1249–52.
168. Taylor LK, Tamblyn R. Reasons for physician nonadherence to electronic drug alerts. *Stud Health Technol Inform* 2004; **107**: 1101–5.
169. 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**: 111–9.
170. Schiff GD, Rucker TD. Computerized prescribing: building the electronic infrastructure for better usage. *JAMA* 1998; **279**: 1024–9.
171. Von Ferber C. Confidentiality Working Group questionnaire. *EURO DURG Bull* 1998; **3**: 3–4.
172. International Society of Pharmacoepidemiology (ISPE). *Guidelines for Good Practice in Data Privacy, Medical Record Confidentiality, and Research in the Interest of Public Health*. http://www.pharmacoepi.org/resources/privacy.cfm (accessed June 2011).

173. Lewis MA. ISPE: present, past and future. *PharmacoEpidemiology Newsletter* 1997; **11**: 4.
174. Guess HA, Goldsmith C, Henry D, Strom BL. Pharmacoepidemiology as a focus for clinical epidemiology in developing countries. *J Clin Epidemiol* 1991; **44** (Suppl): 101–5S.
175. 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 2): I–VIII.
176. Buschiazzo H, Chaves A, Figueras A, Laporte J-R. Drug utilization in Latin America—the example of DURG-LA. *Essential Drugs Monitor* 2003; **32**: 17–8.
177. Ross-Degnan D, Laing R, Quick J, Ali HM, Ofori-Adjei D, Salako L, *et al*. A strategy for promoting improved pharmaceutical use: the International Network for Rational Use of Drugs. *Soc Sci Med* 1992; **35**: 1329–41.
178. Lee D. The International Network for Rational Use of Drugs. *Essential Drugs Monitor* 1997; **24**: 7–8.
179. Dartnell JGA. *Understanding, Influencing and Evaluating Drug Use*. Melbourne, Australia: Therapeutic Guidelines Limited, 2001.
180. Rowe AK, Rowe SY, Vujicic M, Ross-Degnan D,Chalker J, Holloway KA, Peters DH. 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.

CHAPTER 25
Evaluating and Improving Physician Prescribing

Sumit R. Majumdar,[1] Helene Levens Lipton,[2] and Stephen B. Soumerai[3]

[1] Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada

[2] Schools of Medicine and Pharmacy, University of California at San Francisco, San Francisco, CA, USA

[3] Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA

> Research and clinical practice may be on parallel tracks headed in the same direction, but in contact only through rotting ties.
>
> P.P. Morgan, *Are physicians learning what they read in journals?*, 1985.

Introduction

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 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 2010, 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]

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

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 are 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 as all contributing 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 agents are available (e.g., barbiturates instead of benzodiazepines);
- 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., cerebral vasodilators for senile dementia or 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 ACE-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., inhaled corticosteroids for asthma or spironolactone for heart failure);
- failure to prescribe necessary drug therapies (e.g., use of beta-blockers following acute myocardial infarction or use of bisphosphonates after an osteoporotic fracture); and
- failure to achieve recommended therapeutic goals (e.g., systolic blood pressure levels below 140 mmHg or LDL cholesterol levels below 100 mg/dL for the secondary prevention of myocardial infarction).

Specific illustrations of the above problem categories are ubiquitous in the literature. For example, propoxyphene, a toxic and abusable narcotic analgesic, is often prescribed for mild to moderate pain when other safer, more effective analgesics are available.[38,39] 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, 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) health-care system had uncontrolled hypertension (>160/90 mmHg), despite adequate health care 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 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] a 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] and the 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, Light, and Mosteller established that poorly controlled studies produce misleading estimates of the effects of a variety of social programs.[53] Many non-intervention factors can affect medication use over time, such as marketing campaigns, mass media, State or Federal regulatory policies, seasonal effects, changing in staffing of health-care 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 quasiexperimental 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 to be sufficiently valid to merit inclusion within their systematic reviews.[33]

Interrupted time-series designs include multiple observations (often 10 or more) of study populations before and after intervention. Such designs 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 inferences if no simultaneous change in trend is observed for this group.[18,54]

Another popular design that can often lead to interpretable results is the *pre–post with comparison group design*. This design includes a single observation both before and after treatment in a non-

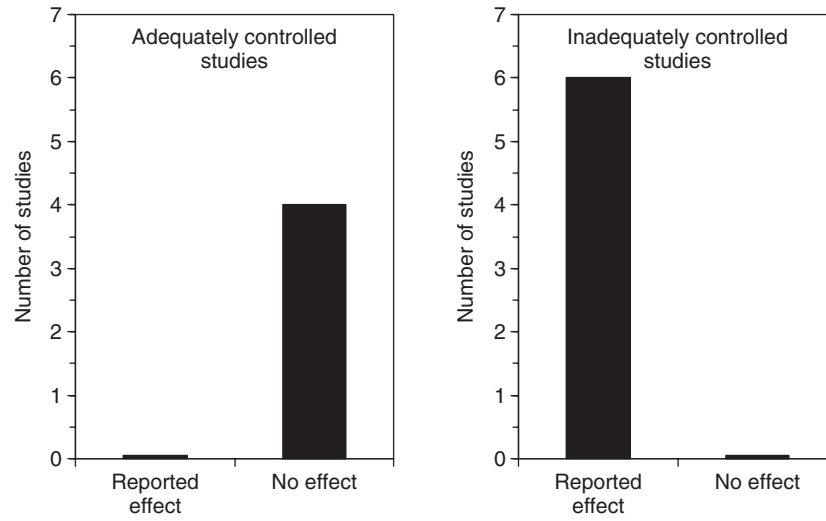


Figure 25.1 Reported effectiveness of dissemination of printed educational materials alone in well-designed versus inadequately controlled studies. Reprinted with permission from the *Milbank Quarterly*.[58]

randomly 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. Although this design controls for many threats to the validity of causal inferences (e.g., due to the effects of testing or maturation), it cannot control for unknown factors (e.g., a regulatory policy) which might result in preintervention differences in trends between study and comparison groups.[53,54]

The weakest, and not uncommon, design is the *one-group, post-only design*, which consists of making only one observation on a single group which 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.[18–24,33] 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.[55–57]

Inadequately controlled studies may exaggerate the effectiveness of many interventions to improve prescribing. For example, as shown in Figure 25.1, inadequately controlled studies of the dissemination of print-only materials used alone (right-hand side) have all reported positive effects on behavior, while well-controlled studies of such strategies (left-hand side) all reported small or non-existent changes in behavior. The "success" of uncontrolled studies is often due to the attribution of pre-existing 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.[57] The results indicated that 50% of the estimated 1-year percent 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,[59] 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.

The above findings 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 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, 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.[53,54]

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 e-mail equivalents) are written to responsible physicians questioning the practice and asking for written responses. Unfortunately, however, the only published research evaluating this method used poorly controlled designs that are unable to control for regression to the mean.[14,17,18] For example, in one often cited DUR study,[59] 50% of prescribing problems were absent several months after letters were sent, suggesting to the non-critical 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 quasiexperiments (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.[60–63] 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.[61–63] These forms of hierarchical "nesting" or statistical "clustering" often lead to accurate point estimates of effect but exaggerated significance levels and inappropriately 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.[61–63] 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.[63]

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.[62–67] Such models allow aggregation at the patient level by accounting for correlation between patients cared for by the same provider or facility. The resulting significance levels for differences in prescribing rates between study and control groups are almost always more conservative (i.e., confidence intervals are "wider") than assuming no intraclass correlation, but are still greater (i.e., confidence intervals "narrower") 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 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 the above 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.[66] 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 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 policy makers 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.[68] What is important is to demonstrate that such interventions are safe, efficacious, and cost-effective *before* widespread adoption. Even a safe and non-efficacious intervention is associated with opportunity costs; if this given

intervention is widely adopted or legislatively mandated, many resources will have been diverted away from other parts of the health-care 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 quasiexperimental designs such as interrupted time-series analysis or staged implementation in which the control population (or regions) receive the intervention after comparative data have been collected.[29,54,62,67,69,70]

Detecting effects on patient outcomes

While a number of studies have demonstrated positive 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 (quasi)-randomized trial of computerized alerts to improve venous thromboprophylaxis for hospitalized patients.[70] Kulcher *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 for 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.[70] The computerized system more than doubled rates of appropriate prophylaxis (34% vs. 14% for controls), although there was still room for improvement. More important, it led to a clinically important (41% decrease) and statistically significant ($p = 0.001$) reduction in adverse clinical events. This is one of a handful of studies that suggest 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: (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 the above 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,[69–71] 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.[69–71]

Currently available solutions

Conceptual framework

A useful starting point for designing an intervention to improve prescribing is to develop a framework for organizing the clinical and non-clinical factors that could help or impede desired changes in clinical behaviors.[7,8,9,10,72] The Theory of Planned Behavior[9,10] is amenable to developing such a framework as is the PRECEDE model.[72] PRECEDE was developed for adult health education programs by Green and Kreuter,[72] 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 e-mail 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 fully the new behavior. Several reviews of the literature have come to a similar conclusion:[20,34,52,73] 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 which do not measure pre-existing trends in target drug use behaviors (see prior "Methodologic Problems" section), only studies using valid experimental or quasiexperimental research designs (e.g., pre–post with comparison group and time-series designs) are discussed.

Disseminating 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 minimal (and often no) effect on actual prescribing practice.[19–23,52,73]

A study of the effect of warning letters mailed to 200000 physicians who were high prescribers of zomepirac sodium corroborates this previous literature.[39] The warning letters, which alerted these physicians to serious or fatal anaphylaxis associated with use of zomepirac, were not associated with any reduction in its use, especially in the face of stronger face-to-face pharmaceutical industry marketing campaigns which may have counteracted the warning messages.

A distinct subset of educational materials are clinical practice guidelines. 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 how 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.[19–23,74–78] 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 as to provide a reasonable control "intervention" with which to compare more effective interventions or strategies.

In summary, simple dissemination of educational materials does not appear to be effective by

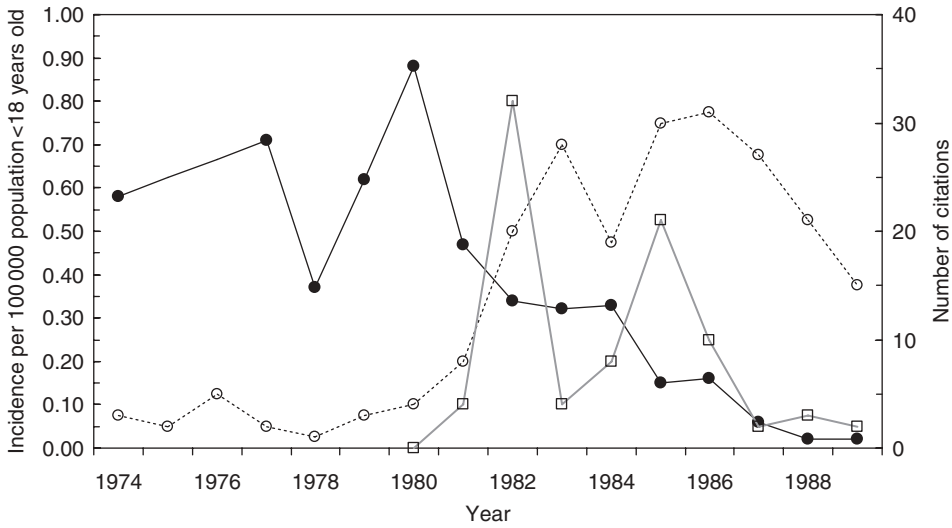


Figure 25.2 Trend in number of (○) medical and (□) lay press citations on aspirin and Reye's syndrome, and the (•) incidence of Reye's syndrome among children. Newspaper index limited to four continuously reporting national newspapers described in text. Reprinted with permission from *Milbank Quarterly*.[82]

itself in altering prescribing patterns, but these materials may provide a necessary *predisposing* foundation for other *enabling* and *reinforcing* strategies.

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. Previous examples include reductions in the use of chloramphenicol (aplastic anemia)[79] calcium channel blockers (myocardial infarction) in response to widespread media warnings.[80] If one considers the prerelease, publication, and intense lay press associated with the results of the Women's Health Initiative RCT a form of multimedia campaign, it is noteworthy that the prescription of estrogen decreased by 38% within 6 months of widespread awareness of the findings of significant harm, and little benefit, associated with the use of hormone therapy.[81]

Figure 25.2 provides data from a US study suggesting that widespread reporting of the risk of Reye's syndrome associated with pediatric aspirin use by the medical and lay press was associated with declines in Reye's syndrome. This media campaign was conducted after Reye's syndrome was linked to aspirin use and antecedent viral illnesses in several epidemiologic studies.[82] 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.

Small group learning

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 non-existent in the literature, especially in non-teaching 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.[83] 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.[19,20,84,85] Even the most rigorous internet-based extensions of traditional continuing medical education had yielded only negligible incremental advantage over more traditional approaches.[85] 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."[84]

Audit and feedback

During the last 25 years, an increasingly 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.[21,23–25] While managers and health policy makers often assume that "feedback" is a unidimensional technique, its many variations have not been well defined or well studied.[25]

One well-studied form of audit and feedback, however, is the patient-level medication profile. 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.[86–88] Likely reasons for the failure of this intuitively appealing approach include: (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.

Other forms of feedback may 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.[21,22,25] 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 entirely supported by a recent methodologically rigorous RCT of the effect of peer-comparison feedback on the prescription of five unrelated groups of medication.[88] 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."[88]

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 important, 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.[88–93]

The level at which feedback is given is another important issue that may differentiate successful and valid programs from questionable ones. For

example, many existing drug utilization review 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,[14,17,18,88] the clinically relevant messages could be unintentionally ignored. 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 program.[14,17,18] For this reason, a more valid method may be to compare patterns of prescribing by individual physicians with clinical guidelines or other more appropriate benchmarks.[90] Lastly, beyond the medium and the message, 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.[89,90]

One advance in the area of audit and feedback, one that attempts to address many of the aforementioned problems, is the development of the concept of the "achievable benchmarks of care" by Kiefe *et al*.[90] The underlying theory is that viewing one's personal performance within the context of peers' performance should be a powerful motivator for change.[90] In essence, the achievable benchmark represents the average performance of the top 10% of local physicians being assessed.[90] 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-randomized controlled trial 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.[90]

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 health care/communications system. In these cases, computerized reminder systems have been developed that 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.[56,70,93–95] In general, prospective reminders are more effective than retrospective feedback; however, such systems are effective only as long as the reminders continue. Further, it is likely that such systems are only effective when clinicians are already predisposed to acting in concert with the protocols. Few data are available on the potential for such systems to reduce inappropriate drug prescribing in cases when physicians have strong beliefs in opposition to recommended practice and (for the most part) various reminder systems have only been studied with a few reminders at a time. "Reminder-fatigue" with concurrent bypassing of computer screens or generalized neglect of all alerts is an important possibility that has not been well documented or studied. It is noteworthy that, even with these caveats, computerized reminders 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 to 11%).[95] Although statistically significant, the mag-

nitude of improvements are far smaller than expected and in many cases might not be considered worthwhile.

Finally, few well-controlled studies are available on the potential for such computerized systems to succeed beyond a "secretarial reminder" function, although early work using locally developed (i.e., homegrown) decision support systems at Brigham and Women's Hospital in Boston, LDS Hospital in Salt Lake City, and the Regenstrief Institute in Indianapolis, show some promise in altering physicians' prescribing decisions in more complex areas such as dosage, schedule, suboptimal choices, and prevention of adverse drug events.[94–96] This promise, however, should not be assumed.[56,95–99] In one older but very rigorous study of advanced computer decision support, Eccles *et al*. conducted a cluster-randomized controlled trial of 60 busy primary care practices in the UK.[99] These practices already had electronic records and electronic prescribing. Eccles *et al*. randomized these practices 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 1 year, there were no significant improvements in any one of more than 40 different quality indicators for either condition.[99] Furthermore, there have also been well-documented harms and adverse events induced by various computerized decision support systems.[97,98] 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 Chapter 45 of this book and recent systematic reviews.[56,93,95,100]

Opinion leaders

The role of local opinion leaders in the adoption of new pharmaceutical agents has been well-documented by Coleman, Katz, and Menzel.[2] Their data indicated that after opinion leaders adopted drugs, other less-integrated physicians eventually followed in a classic curve of technology diffusion. In several studies of diffusion of scientific information on treatment of arthritis and the inappropriate use of Cesarean sections,[101,102] 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.[3,66,101–105] In addition to opinion-leader involvement, these interventions generally included 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. 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."[103]

For example, one RCT demonstrated that opinion leaders could be used to improve prescribing in the treatment of acute myocardial infarction.[66] Hospitals in Minnesota ($N = 37$) were randomized to guideline dissemination, performance feedback, and opinion leaders (intervention), or guidelines and feedback alone (controls). Both the unit of randomization and the unit of analysis were the hospital. Clinical and process data were collected for a year before, and a year after, the intervention (which itself lasted about 6 months). The opinion leaders were asked to promote four separate practices, each consistent with national evidence-based guidelines: increased use of aspirin, increased use of beta-blockers, increased use of thrombolytic therapy in elderly patients, and decreased routine use of lidocaine prophylaxis. Compared to controls, the intervention hospitals successfully increased the use of aspirin (absolute median improvement 13%, $p = 0.04$) and beta-blockers (absolute median improvement 31%, $p = 0.02$). However, there was no improvement in the use of thrombolytic therapy, and all hospitals decreased use of lidocaine by about 50%. This latter finding is evidence of a secular trend, a trend more powerful than the intervention itself, and one that

would have been attributed to the intervention if a weaker study design had been employed. Although the recruitment and use of opinion leaders shows great promise in accelerating the adoption of evidence into practice, overall the results of rigorous opinion leader studies have been mixed,[104] and whether or not such interventions are reproducible across diseases and settings,[105] can improve prescribing for multiple conditions outside the hospital setting, and are cost effective, still remains to be determined.

Academic detailing

A growing number of well-controlled studies support the conclusion that programs combining professionally illustrated educational materials with brief face-to-face visits (15–25 minutes) by university-based pharmacists (academic detailers) or physician counselors or peer-leaders are effective in reducing prescribing of contraindicated or marginally effective therapies in primary care settings. Similarly, several controlled studies of direct educational efforts by pharmacists have also documented improvements in targeted prescribing practices.[22,106,107] 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.

Figure 25.3 provides an example of an educational leaflet briefly summarizing the main educational messages concerning the costs and lack of efficacy of propoxyphene that were emphasized in one RCT of a four-state academic detailing

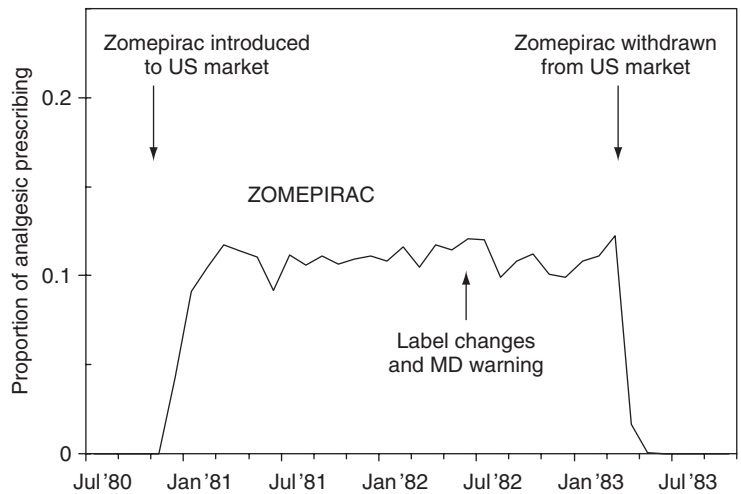


Figure 25.3 Reverse side of graphically illustrated and referenced educational leaflet emphasizing the lack of efficacy and high cost of propoxyphene in comparison to aspirin or acetaminophen. Leaflet was used in face-to-face academic detailing study.[12,75]

program.[75] A formal economic analysis of 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 high benefit-to-cost ratios, even without considering positive spillover effects to non-participating physicians, improved quality of care, or possible cost savings due to elimination of adverse drug effects.[76]

If academic detailing is truly cost neutral (or even cost saving), the main barrier to more widespread use of the strategy is its *perceived* labor intensiveness. Nevertheless, academic detailing is the single most consistently effective method for changing physician practice that has been reported.[19,20,106–108] A number of controlled trials have attempted to replicate the positive results of face-to-face outreach with smaller group outreach sessions, often referred to as "group detailing."[109] Group detailing has the additional advantage of encouraging discussions within the group, which may enhance the diffusion of ideas and increase their impact. For example, in an RCT to improve the treatment of hyperlipidemia in Sweden, Diwan *et al*. randomized 134 health centers.[109] The intervention, for 67 of the health centers, consisted of printed guideline dissemination, an informational video, and four 30-minute group detailing sessions between all health center physicians and a clinical pharmacist, while the control centers only received printed information. Compared to baseline measurements, hyperlipidemia was treated more often for all patients in the intervention centers. Moreover, "first-line prescribing" (use of specific agents in accordance with national guidelines) was 20% higher ($p = 0.03$) for the intervention centers.[109] 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.

Financial incentives and penalties

Although there are few studies of the effects of financial incentives or penalties on physician prescribing behavior *per se*, numerous observational studies suggest that differing payment methods do affect the way that physicians practice medicine.[26–28,110–112] As a general rule, it has also been observed that financial incentives are consistently more powerful than penalties when it comes to changing behavior.

In response to escalating drug costs, an increasing number of physician organizations in the US are entering into capitated drug risk-sharing arrangements with managed care organizations.[22–24,111] 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 must be 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 must be timely and provide specific advice about costs, risks, and possible substitutions; and (iii) physicians must understand and be sensitive to differences in drug pricing. Because it is unlikely that these assumptions (in general) can be 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.[112,113] The most mature pay for performance program is that linked to the UK Quality and Outcomes Framework for primary care.[112] There were more than 100 indicators introduced, and by most standards the incentives were considered generous—up to an additional 25% of an individual physicians' income could be generated within the scheme, and collectively about $1 billion dollars annually was spent by the National Health Service. Early studies, using time-series methods, demonstrated small to modest improvements in some but not all 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.[112] At present, any strategy employing financial incentives to change prescribing should be considered an experiment, and it ought to be studied rigorously, with particular attention to the quality of prescribing, unintended consequences (e.g. cost-shifting), and impacts on patient-related outcomes.

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 inclusion of multiple strategies that predispose, enable, and reinforce 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 it is used at all*) that incorporates 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; and
- 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 so 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 computerized decision support systems safe and effective, and, if so, are they worth the time, effort, and opportunity costs necessary to implement and use them?

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.[22,108] Several formal economic analyses of academic detailing trials demonstrated that the interventions led to a net saving from the societal or organizational perspective.[75,108,114] This is a clear illustration of what Eddy described as "getting more for less," the potential to improve quality and reduce costs simultaneously.[115] 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.[108]

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 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.[116] Less is known about the ambulatory setting,[117–119] 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.[119] 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 relationship between interventions and clinical outcomes would advance the field. While policy-induced reductions in use of essential medications have been associated with adverse events,[120] 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 study. Further, the promise of a comprehensive electronic health record, one that is knowledge-generating and linked to prescriptions, clinical and laboratory information, and claims data, has yet to be fully realized Once there is widespread adoption of these technologies, the fields of health services research and pharmacoepidemiology will enter a new era when innovative measures to improve the quality of prescribing will be implemented and evaluated with heretofore unknown methodologic rigor.

References

1. Morgan PP. Are physicians learning what they read in journals? *Can Med Assoc J* 1985; **133**: 263.
2. Coleman J, 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**: 1271–4.
7. Davis D, Evans M, Jadad A, Perrier L, Rath D, Ryan D, *et al*. The case for knowledge translation: shortening the journey from evidence to effect. *BMJ* 2003; **327**: 33–5.
8. Cabana MD, Rand CS, Power NR, Wu AW, Wilson MH, Abboud PC, *et al*. Why don't physicians follow

clinical practice guidelines? *JAMA* 1999; **282**: 1458–65.
9. Godin G, Belanger AG, Eccles M, Grimshaw J. Healthcare professionals' intentions and behaviors: a systematic review of studies based on social cognitive theories. *Implement Sci* 2008; **3**: 36–46.
10. Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *J Clin Epidemiol* 2005; **58**: 107–112.
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**: 75–85.
12. Soumerai SB, Avorn JL. Principles of educational outreach ("academic detailing") to improve clinical decision making. *JAMA* 1990; **263**: 549–56.
13. Auerbach AD, Landefield CS, Shojania KG. The tension between needing to improve care and knowing how to do it. *N Engl J Med* 2007; **357**: 608–13.
14. Hennessy S, Soumerai SB, Lipton HL, Strom BL. Drug utilization review. In: Strom BL, ed. *Pharmacoepidemiology*, 4th edn. Chichester: John Wiley and Sons, 2005, pp. 439–54.
15. Lipton HL, Byrns PH, Soumerai SB, Chrischilles EA. Pharmacists as agents of change for rational drug therapy. *Int J Technol Assess Health Care* 1995; **11**: 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 of Systematic Reviews). In: *The Cochrane Collaboration*, Issue 1. John Wiley & Sons, 2009.
17. Soumerai SB, Lipton HL. Computer based drug utilization review—risk, benefit, or boondoggle? *N Engl J Med* 1995; **332**: 1641–4.
18. Hennessy S, Bilker WB, Zhou L, Weber AL, Brensinger C, Wang , *et al*. Retrospective drug utilization review, prescribing errors, and clinical outcomes. *JAMA* 2003; **290**: 1494–9.
19. Forsetlund L, Bjorndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, *et al*. Continuing education meetings and workshops: effects on professional practice and health care outcomes (Cochrane Database of Systematic Reviews). In: *The Cochrane Collaboration*, Issue 2. John Wiley & Sons, 2009.
20. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, *et al*. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care* 2001; **39** (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**: 584–90.
22. Pearson SA, Ross-Degnan D, Payson A, Soumerai SB. Changing medication use in managed care: a critical review of the available evidence. *Am J Manag Care* 2003; **9**: 715–31.
23. Shojania KG, Grimshaw JM. Evidence based quality improvement: state of the science. *Health Aff* 2005; **24**: 138–150.
24. Mattke S, Seid M, Ma S. Evidence for the effect of disease management: is one billion dollars a year a good investment? *Am J Manag Care* 2007; **13**: 670–6.
25. Jamtvedt G, Young JM, Kristoffersen DT, Obrien 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**: 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**: 2224–32.
27. Bloor K, Freemantle N. Lessons from international experience in controlling pharmaceutical expenditure II: influencing doctors. *BMJ* 1996; **312**: 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**: 217–52.
30. Soumerai SB, McLaughlin TJ, Ross-Degnan D, Casteris CS, Bollini P. Effects of 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**: 650–6.
31. Smalley WE, Griffin MR, Fought RL, Sullivan L, Ray WA. Effect of a prior authorization requirement on the use of nonsteroidal anti-inflammatory drugs by Medicaid patients. *N Engl J Med* 1995; **332**: 1612–17.
32. Tamblyn R, Laprise R, Hanley JA, Abrahamowicz M, Scott S, Mayo N, *et al*. Adverse events associated

with prescription drug cost-sharing among poor and elderly persons. *JAMA* 2001; **285**: 421–9.
33. Cochrane Effective Practice and Organization of Care Group. http://epoc.cochrane.org/ (accessed Sept. 2010).
34. Spinewine A, Schmader KE, Barber N, Hughes C, Lapane KL, Swine C, Hanlon JT. Appropriate prescribing in the elderly: how well can it be measured and optimized. *Lancet* 2007; **370**: 173–84.
35. Majumdar SR, Soumerai SB. Why most interventions to improve physician prescribing do not seem to work. *Can Med Assoc J* 2003; **169**: 30–1.
36. McGlynn EA, Asch SM, Adams J, Kessey J, Hicks J, DeCristofaro A, *et al*. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003; **348**: 2635–45.
37. Lee TH. A broader concept of medical errors. *N Engl J Med* 2002; **347**: 1965–7.
38. Soumerai SB, Avorn J, Gortmaker S, Hawley S. Effect of government and commercial warnings on reducing prescription misuse: the case of propoxyphene. *Am J Public Health* 1987; **77**: 1518–23.
39. Ross-Degnan D, Soumerai SB, Fortess EE, Gurwitz J. Examining product risk in context: market withdrawal of zomepirac sodium as a case study. *JAMA* 1993; **270**: 1937–42.
40. Gonzales R, Steiner JF, Lum A, Barrett PH. Decreasing antibiotic use in ambulatory practice: impact of a multidimensional intervention on the treatment of uncomplicated acute bronchitis in adults. *JAMA* 1999; **281**: 1512–19.
41. Gurwitz JH, Field TS, Rochon P, Judge J, Harrold LR, Bell CM, *et al*. Effect of computerized physician order entry with clinical decision support on adverse drug events in the long term care setting. *J Am Geriatr Soc* 2008; **56**: 2225–33.
42. Gurwitz JH, Field TS, Harrold LR, Rothschild J, Debellis K, Seger AC, *et al*. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *JAMA* 2003; **289**: 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**: 1957–63.
44. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; **89**: 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**: 467–72.
46. Soumerai SB, Avorn J, Ross-Degnan D, Gortmaker S. Payment restrictions for prescription drugs in Medicaid: effects on therapy, cost, and equity. *N Engl J Med* 1987; **317**: 550–6.
47. Soumerai SB, Ross-Degnan D, Avorn J, McLaughlin TJ, Choodnovskiy I. Effects of Medicaid drug-payment limits on admission to hospitals and nursing homes. *N Engl J Med* 1991; **325**: 1072–7.
48. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectibility of man. *N Engl J Med* 1976; **295**: 1351–5.
49. Kravitz RL, Epstein RM, Feldman MD, Franz CE, Azari R, Wilkes MS, *et al*. Influence of patients' requests for direct to consumer advertised antidepressants: a randomized controlled trial. *JAMA* 2005; **293**: 1995–2002
50. Schwartz RK, Soumerai SB, Avorn J. Physician motivations for non-scientific drug prescribing. *Soc Sci Med* 1989; **28**: 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**: 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**: 1738–42.
53. Gilbert JR, Light RJ, Mosteller F. Assessing social innovations: an empirical base for policy. In: Bennett CH, Lumsdaine AA, eds. *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*. New York: Academic Press, 1975.
54. Cook TD, Campbell DT. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin, 1979.
55. Shojania KG, Duncan GW, McDonald KM, Wachter RM. Safe but sound: Patient safety meets evidence-based medicine. *JAMA* 2002; **288**: 508–13.
56. Chaudry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, *et al*. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006; **144**: 742–52.
57. Soumerai SB, Ross-Degnan D, Gortmaker S, Avorn J. Withdrawing payment for non-scientific drug therapy: intended and unexpected effects of a large-scale natural experiment. *JAMA* 1990; **263**: 831–9.
58. Soumerai SB, McLaughlin TJ, Avorn J. Improving drug prescribing in primary care: a critical analysis of the experimental literature. *Milbank Q* 1989; **67**: 268–317.

59. Groves R. Therapeutic drug-use review for the Florida Medicaid Program. *Am J Hosp Pharm* 1985; **42**: 316–19
60. Donner A, Birkett N, Buck C. Randomization by cluster—samples size requirements and analysis. *Am J Epidemiology* 1981; **114**: 906–14.
61. Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med* 2001; **20**: 391–9.
62. Freemantle N, Eccles M, Wood J, Mason J, Nazareth I, Duggan C, *et al*. A randomized trial of evidence-based outreach: rationale and design. *Control Clin Trials* 1999; **20**: 479–92.
63. Divine GW, Brown JT, Frazier LM. The unit of analysis error in studies about physicians' patient care behavior. *J Gen Intern Med* 1992; **7**: 623–9.
64. Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford: Clarendon Press, 1996.
65. McCullagh P, Nelder JA. *Generalized Linear Models*, 2nd edn. London: Chapman & Hall, 1989.
66. Soumerai SB, McLaughlin TJ, Gurwitz JH, Guadagnoli E, Hauptman PJ, Borbas C, *et al*. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *JAMA* 1998; **279**: 1358–63.
67. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methods* 2006; **6**: 54–63.
68. The Leapfrog Group for Patient Safety. http://www.leapfroggroup.org/ (accessed Sept. 2010).
69. Peterson ED, Roe MT, Mulgund J, DeLong ER, Lytle BL, *et al*. Association between hospital process performance and outcomes among patients with acute coronary syndromes. *JAMA* 2006; **295**: 1912 –20.
70. Kulcher N, Koo S, Quiroz R, Cooper JM, Paterno MD, Soukonnikov B, Goldhaber SZ. Electronic alerts to prevent venous thromboembolism among hospitalized patients. *N Engl J Med* 2005; **352**: 969 –77.
71. Brook RH, McGlynn EA. Measuring quality of care. *N Engl J Med* 1996; **335**: 966–9.
72. Green LW, Kreuter MW. *Health Promotion Planning: An Educational and Environmental Approach*. Mountain View, CA: HJ Kaiser Foundation, 1991.
73. Farmer AP, Legare F, Turcot L, Grimshaw J, Harvey E, McGowan JL, Wolf F. Printed educational materials: effect on professional practice and healthcare outcomes. (Cochrane Database of Systematic Reviews). In: *The Cochrane Collaboration*, Issue 2. John Wiley & Sons, 2009.
74. Schaffner W, Ray WA, Federspiel CF, Miller WO. Improving antibiotic prescribing in office practice: a controlled trial of three educational methods. *JAMA* 1983; **250**: 1728–32.
75. Avorn JL, 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.
76. Soumerai SB, Avorn JL. Economic and policy analysis of university-based drug "detailing." *Med Care* 1986; **24**: 313–31.
77. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; **342**: 1317 –22.
78. Lomas J, Anderson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice? The effect of consensus statements on the practice of physicians. *N Engl J Med* 1989; **321**: 1306–11.
79. Wade OL, Hood H. Prescribing of drugs reported to cause adverse reactions. *Br J Prev Soc Med* 1972; **26**: 205–11.
80. Maclure M, Dormuth C, Naumann T, McCormack J, Rangno R, Whiteside C, *et al*. Influences of educational interventions and adverse news about calcium channel blockers on first line prescribing of antihypertensive drugs to elderly people in British Columbia. *Lancet* 1998; **352**: 943–8.
81. Majumdar SR, Elmasi EA, Stafford RS. Promotion and prescribing of hormone therapy after report of harm by the Women's Health Initiative. *JAMA* 2004; **292**: 1983–8.
82. 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.
83. Inui TS, Yourtee EL, Williamson JW. Improved outcomes in hypertension after physician tutorials: a controlled trial. *Ann Intern Med* 1976; **84**: 646–51.
84. 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**: 511–15.
85. Fordis M, King JE, Ballantyne CM, Jones PH, Schneider KH, Spann SJ, *et al*. Comparison of the instructional efficacy of internet-based CME with live interactive CME workshops: a randomized controlled trial. *JAMA* 2005; **294**: 1043–51.
86. Hershey CO, Porter DK, Breslau D, Cohen DI. Influence of simple computerized feedback on pre-

scription changes in an ambulatory clinic: a randomized clinical trial. *Med Care* 1986; **24**: 472–81.
87. Koepsell TD, Gurtel A, Diehr PH, Temkin NR, Helfand KH, Gleser MA, *et al*. The Seattle evaluation of computerized drug profiles: effects on prescribing practices and resource use. *Am J Public Health* 1983; **73**: 850–5.
88. O'Connell DL, Henry D, Tomlins R. Randomized controlled trial of effect of feedback on general practitioners' prescribing in Australia. *BMJ* 1999; **318**: 507–11.
89. 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**: 250–6.
90. 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**: 2871–9.
91. 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**: 88–93.
92. Gehlbach SH, Wilkinson WE, Hammond WE, Clapp NE, Finn AL, Taylor WJ, *et al*. Improving drug prescribing in a primary care practice. *Med Care* 1984; **22**: 193–201.
93. 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**: 1409–16
94. 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**: 965–70.
95. 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**: 216 –22.
96. Bates DW, Gawande AA. Improving patient safety with information technology. *N Engl J Med* 2003; **348**: 2256–34.
97. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, Strom BL. Role of computerized physician order entry systems in facilitating medication errors. *JAMA* 2005; **293**: 1197–203.
98. Han YY, Carcillo JA, Venkataraman ST, Clark RS, Watson RS, Nguyen TC, *et al*. Unexpected increased mortality after implementation of a commercially sold computerized physicians order entry system. *Pediatrics* 2005; **116**: 1506–12.
99. Eccles M, McColl E, Steen N, Rousseau N, Grimshaw J, Parkin D, *et al*. Effect of computerized evidence based guidelines on management of asthma and angina in adults in primary care: a cluster randomized controlled trial. *BMJ* 2002; **325**: 941–8.
100. Garg AX, Adhikari NK, McDonald H, Rosas-Arrellano MP, Devereaux PJ, Beyene J, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005; **293**: 1223–8.
101. Stross JK, Bole GG. Evaluation of a continuing education program in rheumatoid arthritis. *Arthritis Rheum* 1980; **23**: 846–9.
102. Lomas J, Enkin M, Anderson G, Hannah WJ, Vayda E, Singer J. Opinion leaders vs. audit and feedback to implement practice guidelines. Delivery after previous cesarean section. *JAMA* 1991; **265**: 2202–7.
103. 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.
104. Doumit G, Gattellari M, Grimshaw J, O'Brien MA. Local opinion leaders: effects on professional practice and health care outcomes (Cochrane Database of Systematic Reviews). In: *The Cochrane Collaboration,* Issue 1. John Wiley & Sons, 2009.
105. 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**: 22–9.
106. Freemantle N, Nazareth I, Eccles M, Wood J, Haines A, Evidence-based OutReach trialists. A randomized controlled trial of the effect of educational outreach by community pharmacists on prescribing in UK general practice. *Br J Gen Practice* 2002; **52**: 290–5.
107. O'Brien MA, Rogers S, Jamtvedt G, Oxman AD, Odgaard-Jensen J, Kristoffersen DT, *et al*. Educational outreach visits: effects on professional practice and health care outcomes (Cochrane Database of Systematic Reviews). In: *The Cochrane Collaboration*, Issue 4. John Wiley & Sons, 2008.
108. Mason J, Freemantle N, Nazareth I, Eccles M, Haines A, Drummond M. When is it cost effective to change behavior of health professionals? *JAMA* 2001; **286**: 2988–92.

109. 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**: 705–11.
110. Gosden T, Forland F, Kristiansen IS, Sutton M, Leese B, Giuffrida A, *et al*. Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians (Cochrane Database of Systematic Reviews). In: *The Cochrane Collaboration*, Issue 2, 2008. John Wiley & Sons.
111. Lipton HL, Kreling DK, Collins T, Hertz KC. Pharmacy benefit management companies: dimensions of performance. *Ann Rev Public Health* 1999; **20**: 361–401.
112. 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**: 368–78.
113. Pearson SD, Schneider EC, Kleinman KP, Coltin KL, Singer JA. The impact of pay for performance on healthcare quality in Massachusetts, 2001–2003. *Health Affairs* 2008; **27**: 1167–76.
114. Simon SR, Rodriguez HP, Majumdar SR, Kleinman KP, Warner C, Salem-Schatz S, *et al*. Economic analysis of a randomized trial of academic detailing interventions to improve use of antihypertensive medications. *J Clin Hypertens* 2007; **9**: 15–20.
115. Eddy D. Rationing resources while improving quality: how to get more for less. *JAMA* 1994; **272**: 817–24.
116. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, *et al*. Incidence of adverse drug events and potential adverse drug events: implications for prevention. *JAMA* 1995; **274**: 29–34.
117. Gandhi TK, Weingart SN, Borus J, Seger AC, Peterson J, Burdick E, *et al*. Adverse drug events in ambulatory care. *N Engl J Med* 2003; **348**: 1556–64.
118. Tierney WM. Adverse outpatient drug events—a problem and an opportunity. *N Engl J Med* 2003; **348**: 1587–9.
119. 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**: 2414–20.
120. Majumdar SR, Soumerai SB. The unhealthy state of health policy research. *Health Aff (Millwood)* 2009; **28**: 900–8.

CHAPTER 26
Pharmacoepidemiologic Studies of Vaccine Safety

Robert T. Chen,[1] Jason M. Glanz,[2] and Claudia Vellozzi[3]

[1] Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention Atlanta, GA, USA

[2] Institute for Health Research, Kaiser Permanente Colorado, and Department of Epidemiology, Colorado School of Public Health, Denver, CO, USA

[3] Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA

Introduction

Vaccines are among the most cost effective and prevalent public health interventions.[1,2] Where immunization is 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 (AEFIs) 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 resurgence of pertussis.[12] Similar concerns in the United States led to lawsuits, substantial vaccine price increases, loss of vaccine manufacturers,[13] and were a potential deterrent to the development of new vaccines.[14] More recently, concerns about the safety of mercury-based thimerosal preservative used in vaccines[15,16] and the safety of anthrax[17] and smallpox vaccines[18,19] have affected the stability of US civilian and military immunization programs, respectively. In the United Kingdom, a case series report of autism following 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.[20,21] Similarly, vaccine safety concerns have affected public acceptance of hepatitis B vaccine in France,[22] oral polio vaccine (OPV) in Nigeria,[23,24] and 2009 H1N1 vaccine in several countries.[25,26]

In the early 1990s, a review by the Institute of Medicine (IOM) in the United States[27,28] noted that vaccine safety knowledge and research capacity had been limited by several factors: (i) inadequate understanding of biologic mechanisms underlying adverse events; (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 concluded that, "if research capacity and accomplishments [are] not improved, future reviews of vaccine safety [will be] similarly handicapped."

Pharmacoepidemiology has played a vital role in providing the scientific methods for assessing vaccine safety in the United States,[29] Europe,[30] and globally.[11,31–33] Many research and knowledge gaps continue to be identified in each IOM review of specific immunization safety controversies since

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

2001, ranging from autism to unexpected infant deaths.[17,34–40a] In this chapter, we discuss the major differences in how epidemiology is applied to vaccines and other pharmaceutical products, giving consideration to both policy and methodology.

Clinical problems to be addressed using pharmacoepidemiologic research

Policy issues

Vaccines share many characteristics with other pharmaceuticals, such as their phased development and licensure, but differ fundamentally in many ways. Understanding these differences is important to appreciate the policy context of vaccine safety and the role of pharmacoepidemiology. Vaccines, for example, are biological products that are inherently more complex than most small-molecule drugs in terms of both constituent components and the production process.[41,42] Each component of the vaccine formulation—the immunogen, conjugated protein,[43] preservative,[15] adjuvant,[44,45] stabilizer,[46–49] diluent,[50] and other excipients—has its respective safety considerations (e.g., sourcing, production, quality assurance, safety profile), individually as well as combined.[51] Programmatic errors such as mixing up vaccine vials and administration errors such as unsafe injections can also be a concern, especially in low-income countries.[50,52]

A higher standard of safety is also expected of vaccines. In contrast to pharmaceuticals, most of which are administered to people 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 substantially lower than to products administered to people who are already sick. This lower risk tolerance for vaccines translates into a need to investigate the possible causes of much rarer adverse events following vaccinations than would be acceptable for other pharmaceuticals. Events that occur at approximately 1/100 000 to 1/1 000 000 doses, such as acute encephalopathy after whole-cell pertussis vaccine,[27,53] Guillain–Barré syndrome (GBS) after swine[54] or 2009 H1N1[55] influenza vaccines, and oral polio vaccine-associated paralytic polio (VAPP),[56] 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.[57]

The cost and the difficulty of studying events increase with their rarity, however (see Chapter 3). Furthermore, the ability to provide definitive conclusions from epidemiologic studies of rare events is poor. Attributable risks in the order of 1/10 000–1/100 000 are on the margin of resolution for epidemiologic methods.[27,58] This challenge was illustrated during the whole-cell pertussis vaccine safety concern in the late 1970s.[12] All British children 2 to 35 months of age hospitalized for several neurological illnesses over 3 years (n = 1167) were enrolled in a very large case–control study.[59] The finding of a significant association between vaccine and permanent brain damage was based on only seven exposed cases.[53] Whether or not this study finding was valid, it generated much controversy in and out of the courts.[27,60] Interestingly, a recent study suggests a possible *de novo* genetic mutation predisposing risk factor for the cases.[61]

Despite considerably more robust data linking GBS with the swine influenza vaccine,[54] subsequent controversy[62,63] resulted in a court-ordered independent re-examination of the data[64] and ultimately to a partial repetition of the study, confirming the initial findings.[65] Robust results from two studies on rhesus rotavirus vaccine and intussusception[66,67] have also been challenged.[68,69]

Perhaps not surprisingly, but adding to the confusion, much of the published literature on vaccine safety 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[20]) rather than controlled studies with adequate statistical power.[27,28] This problem has been ameliorated recently with the advent of carefully controlled, large, linked database studies in the United States, United Kingdom, and Denmark.[33,70]

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 do so by law or regulation for public health reasons.[71] Such requirements have been implemented by public health authorities because many VPD (e.g., measles) are highly contagious. 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 risk–benefit ratio diverges from the societal risk–benefit 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,[75,76] as was experienced in the United Kingdom with pertussis[12] and measles.[21] A similar policy consideration occurs for some mandatory military vaccinations such as anthrax[17] and smallpox,[18] where a higher vaccine reaction rate may be accepted in exchange for battlefield 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 was administered to about 400 000 people in the "Cutter Incident", resulting in 260 polio cases.[77,78] The following incidents 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" such as immunizations: (i) polio vaccine contaminated by simian virus 40 may have been received by millions of people during the 1950s;[79] (ii) some vaccines may have contained gelatin stabilizers produced in cattle infected with bovine spongiform encephalopathy;[80] (iii) some US children were exposed to high levels of ethyl mercury from thimerosal preservatives in vaccines;[15] and (iv) two of the new rotavirus vaccines were contaminated by a porcine circo virus.[81] These concerns are the basis for strict regulatory control and other oversight of vaccines by National regulatory authorities such as the Food and Drug Administration (FDA), European Medicines Agency (EMA),[41] and the World Health Organization (WHO).[82] Modern technology will continue to improve the ability to detect contaminants in vaccines and influence regulatory decisions during manufacturing; postlicensure monitoring will continue to be important should such findings raise safety concerns.[81]

Very high standards of accuracy and timeliness 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 (oral and inactivated poliovirus vaccines being the best known exception). The decision to withdraw a vaccine[62] or switch between strains may also have wide ramifications.[56,83] In 1992, the United Kingdom withdrew the license of mumps vaccines containing the Urabe strain after studies suggested a high rate of vaccine-associated meningitis.[84] The manufacturers subsequently withdrew this product worldwide, leaving countries without an alternative vaccine if the Urabe strain was the sole mumps vaccine licensed.[85,86] Safety concerns led to the withdrawal in the early 2000s of what were then the only licensed vaccines against rotavirus[68] and Lyme disease,[87] rendering these vaccines unavailable anywhere. Establishing associations of adverse events with vaccines and timely measurement of the attributable risk are critical in placing adverse events in the proper risk–benefit perspective. An erroneous association or attributable risk, especially with misinformed media or websites,[88,89] can undermine confidence in a vaccine and have disastrous consequences for vaccine acceptance and disease incidence.[21] 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.[90,91]

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.[92] Accurate assessment of whether adverse events can be caused by specific vaccines is essential to a fair and efficient vaccine injury compensation program.[93] In the United States, for example, the Vaccine Injury Table contains the vaccines, adverse events, and intervals after which no-fault decisions are made in favor of the claimants.[92] Periodic revisions of the Vaccine Injury Table are necessary to reflect the best scientific information on associations between vaccines and adverse events, especially following introduction of new vaccines.[94]

Finally, recommendations for 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.[95] When the target diseases are close to eradication, high vaccine complication rates relative to that of the target wild disease may lead to discontinuation or decreased use of the vaccine, as was done with smallpox vaccine.[96] Another example was oral polio vaccines, where there was a shift to either inactivated polio vaccine (IPV)[83,97] or sequential IPV/oral polio vaccine (OPV),[56] in order to control OPV-associated paralytic polio and circulation of OPV-derived polio virus.[98] 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.[99]

With the renewed fears of bioterrorism, stopping immunizations and allowing formation of lacunae in herd immunity no longer seems advisable.[32,100] Almost all immunizations will therefore 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 health-care providers (let alone parents) have not ever seen a case of the wild VPD. Each future generation must therefore be convinced of the need to be immunized despite an increasingly ancient experience of wild disease but contemporary fear of vaccine adverse events.

Research in vaccine safety—while applying pharmacoepidemiologic principles—can help to distinguish true vaccine reactions from coincidental events,[33,101,102] estimate their attributable risk,[53,54,67,103–107] identify risk factors that may permit development of valid contraindications,[53,108] and, if the pathophysiologic mechanism becomes known, develop safer vaccines.[109,110] 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[111] (see Chapter 1). Phase I trials usually number their subjects in the tens and can only detect extremely common adverse events. 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.[112] Such studies can also affect the choice of the candidate vaccine for Phase III.[113]

Sample sizes for Phase III vaccine trials generally range between 5000 and 10 000 people, which is larger than most drug trials. In extremis, more than 600 000 schoolchildren were enrolled in the famous Francis field trial of inactivated Salk poliovirus vaccine.[114] To help rule out links with a rarer outcome such as intussusceptions (background rate ~ 5 per 1000 infant years), the second generation rotavirus vaccine trials enrolled approximately 70 000 infants.[115,116] 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 <30 days).[113] 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 adverse event with the vaccine are relatively straightforward.[27,28] Brazilian investigators also used such a design to compare the risk of aseptic meningitis among three mumps vaccine strains.[117]

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, below). In the Phase III trials for infant DTaP, a standard case definition was developed for efficacy, but ironically not for safety—the main reason for the development of DTaP.[118] For example, definitions of high fever across trials varied by the temperature (39.5° versus 40.5°C), the mode of measurement (oral versus rectal), and time after vaccination measured (48 versus 72 hours).[119] 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.[120]

The finding of delayed excess mortality in some recipients of high-titer measles vaccine in developing countries,[121] now believed by some to be due to a change in vaccine sequence[122] or non-specific effects of vaccinations,[123] has also led to a call for increasing the current limited duration of follow-up for AEFI in most trials.[124,125] Furthermore, many of the 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 therefore need to be done in settings where the pharmacovigilance infrastructure is limited or non-existent.[126,127]

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 under-respond and others over-respond to an immunization with respect to immunogenicity and reactogenicity. 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 were at risk of being unfairly labeled as "antivaccine" if they questioned the safety of receipt of subsequent doses.

Despite over 200 years since Jenner first 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 logistical as scientific. Modern medicine cannot make progress on rare disorders such as leukemia (or rare serious vaccine reactions) by relying on primary care providers alone. Instead, tertiary subspecialties with adequate referral base and research funds (e.g., hematology/oncology) are needed. With the exception of certain regions in Italy,[128] Australia,[129,130] and six civilian Clinical Immunization Safety Assessment (CISA) sites[130a] and four military Vaccine Research Centers (VHC) in the United States,[131] 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 vaccine adverse event syndromes.[20,132] The diversity of vaccine exposures (active/ passive, live/ killed, 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. But most importantly, such CISA-type centers could potentially prevent outliers (for example, based on genetic susceptibility) in reactogenicity response from becoming antivaccine activists by recruiting them into "win–win" opportunity to improve our scientific understanding and hopefully eventual prevention of vaccine reactions, as done recently in a clinical trial of safer booster dose

in children with extensive limb swelling after pertussis vaccination.[133]

Methodologic problems to be addressed using pharmacoepidemiologic research

Signal detection

Because biologics such as vaccines are generally manufactured in living systems rather than through chemical synthesis as for drugs, variation in rate of adverse reactions by manufacturer or even lot might be expected.[134–136] Surveillance systems need to detect such potential aberrations in the expected number and type of adverse events in a timely manner. Some factors make identification of true 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 adverse event rates can be calculated, it may be difficult or impossible to know if an aberration has occurred.[137] 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 most public health surveillance systems, which focus on either a single exposure (e.g., lead) or single disease outcome (e.g., measles), vaccine safety surveillance systems need to examine multiple exposures (e.g., different vaccine antigens, manufacturers, 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[138] as due to data analysis.[139] The trade-off 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 adverse events in the same system, however.[140] Nevertheless, the goal of early detection of an aberrant cluster of new adverse events 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.[141] The sensitivity of surveillance may be lower and the difficulty of assessing misclassification greater, however. Alternatively, if the reporting form is open-ended,[142] this may increase the sensitivity of surveillance but only at the cost of sorting through many non-specific reports. Definitions can then be applied at the time of analysis. But substantial variation in diagnostic workup and description of events makes *post hoc* classification difficult without additional follow-up information, which in turn is usually costly.

Historically, it was difficult 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 Classifications and Case Definitions, below) is addressing this gap.[143]

Assessment of causality

Aside from events such as local reaction or anaphylaxis, assessing whether any adverse event 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[19]), or repeat exposures resulting in the same adverse event (e.g., alopecia and hepatitis B vaccination[138]), or a vaccine-specific laboratory finding (e.g., Urabe mumps vaccine virus isolation[144]) 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 studies are 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 older people may be particularly difficult. In the United States, 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 potential confusing maze of vaccination history misclassifications.[137,145]

For example, even though only the acellular pertussis vaccine is available in the United States, adverse event reports of the old whole-cell pertussis "DTP" vaccine continue to be received—presumably due to errors in recording by immunization providers due to old habits. 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 such as polio or hepatitis B vaccines are administered simultaneously, at different dose series in the schedule, at different ages, using different lots of vaccine, and the number of permutations of vaccine exposures that need assessment for potential safety concerns quickly becomes formidable.[146] The near-unique availability of complete documentation of vaccine exposure on a large cohort of children in the Vaccine Safety Datalink (VSD) project allowed the evaluation of the safety of thimerosal preservatives.[147,148]

Outcome

Because of the higher standard required for vaccine safety (as discussed previously), events being assessed are frequently extremely 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 Institute of Medicine committee that evaluated whether the United Kingdom's National Childhood Encephalopathy Study should be replicated in the United States.[58]

The difficulty in achieving adequate statistical 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 subpopulations with special indications). This challenge is well illustrated in studies of the potential association between GBS, which occurs at a background rate of about one per 100 000 person years, and various vaccines. Study of GBS after newly introduced meningococcal conjugate vaccine (MCV) required assembling data from 9 million adolescents.[149] A retrospective study of GBS after the 1992–1994 influenza vaccinations required assessing hospital records of over 20 million people for 2 years.[150] Recently, active GBS case finding among a population of 45 million may have detected an attributable risk of one additional case of GBS per million H1N1 vaccinations.[55]

Whenever both the rarity of the adverse outcome and the number of exposures limits 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.[49]

Many adverse events hypothesized to be caused by vaccines have poorly defined etiologies (e.g., encephalopathy,[151] GBS,[54] chronic fatigue syndrome,[152] narcolepsy,[152a] sudden infant death syndrome [SIDS][153]) and 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, below). Predicting the onset of illness is critical in calculating the risk interval. For certain hypothesized vaccine adverse events, there is no known biological mechanism to allow prediction of the risk interval. Diseases with insidious or delayed onset like autism,[20] inflammatory bowel disease,[154] and multiple sclerosis[155] 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, and underlying health disorders.[121]

Another challenge is that serious vaccine adverse events 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.[156–159] Such studies can be prohibitively expensive, unless all 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.[66,106,160–162] 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 AEFIs are age-dependent, age may confound exposure–outcome relations (e.g., diphtheria–tetanus–pertussis [DTP] vaccine and febrile seizures or SIDS[163]). Consequently, such factors must be controlled, generally by matching and subsequent adjustment in the statistical analysis.

To address these limitations, various self-controlled study designs have been developed and implemented.[106,164–172] 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.[173,174] For details on these methods see Methodologic Approaches section below.

More difficult to control are factors leading to delayed vaccination or non-vaccination.[121] Such factors (e.g., low socioeconomic status, preceding illness) may confound studies of vaccine adverse events and lead to under-estimates of the true relative risks. The extent of bias introduced by confounding can be examined as a function of six variables (Table 26.1). Relatively little is known about the nature, frequency, and implications of these variables, however.[121]

Table 26.1 Variables determining the extent of bias attributable to confounding in studies of vaccine adverse events (AE)[122]

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.

Currently available solutions

Prelicensure

Whenever potentially important safety signals are detected in prelicensure trials (e.g., intussusceptions after rotavirus vaccine),[175] it is critical that they are pursued postlicensure.[176] 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 have argued that larger experimental trials may be needed to better assess rare but serious vaccine risks.[74,177] This could be done either with larger prelicensure trials, as has been done with antipyretics in children[177–179] and the postrhesus rotavirus vaccine trials,[116] or in some organized step-wise manner postlicensure (e.g., registry of the first million vaccinations), prior to universal recommendation of the vaccine for entire birth cohorts.[178] Even with these measures, separate large-scale, long-term randomized intervention trials would theoretically be the only way to study unforeseen delayed vaccine adverse effects[74] or non-specific effects of immunizations[123] such as those seen with killed[180] or high-titer measles vaccines.[181,182] Such trials would have to overcome major concerns about the ethics of withholding efficacious vaccines from people in need. Therefore, a more likely way forward probably lies in maximizing both the pre-and postlicensure assessment processes as discussed in this Chapter.

In addition to standardized case definitions for adverse outcomes, 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 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 lifecycle.[183]

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 adverse event with the experimental exposure (e.g., 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 prone to bias.[184] In an era of increasing automation of medical records and sophistication of methods for detecting non-random clusters or elevated rates, similar approaches to assessing prelicensure safety data are needed. 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.[126,127,185]

Postlicensure

Passive surveillance or spontaneous reporting systems (SRS)

Informal or formal passive surveillance or spontaneous reporting systems (SRS) have been the cornerstone of most vaccine safety monitoring systems because of their simplicity and relatively low cost.[186] The national reporting of vaccine adverse events can be done through the same reporting channels as those used for other adverse drug reactions,[187] as is the practice in most European countries,[188,189] Japan,[190] and New Zealand.[191] Historically, however, few countries have forwarded their AEFI reports to the Uppsala Monitoring Center[192] (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,[193] Brazil,[194] Canada,[195] Cuba,[196] Denmark,[197] India,[198] Italy,[199] Germany,[200] Mexico,[186] Netherlands,[201] New Zealand,[202] Switzerland,[203] and the United States.[142] Vaccine manufacturers also maintain SRS for their products,[204] which are usually forwarded subsequently to appropriate national regulatory authorities.[41]

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 United States, 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 vaccine adverse events commonly are first reported by the health-care providers who administered the vaccine. In many countries, such health-care providers also participate in public health surveillance for other diseases. Public health authorities (e.g., Centers for Disease Control, CDC) therefore commonly lead or collaborate with the vaccine licensure and regulatory agency (e.g., the US FDA) in developing AEFI reporting systems. A similar model for harmonization and avoiding duplication is followed in Canada[205] and six European countries,[189] and is highly recommended for other countries.[206]

The US experience

The US National Childhood Vaccine Injury Act of 1986 mandated for the first time that health-care providers report certain adverse events after immunizations (Table 26.2).[207] The Vaccine Adverse Event Reporting System (VAERS) was implemented jointly by the CDC and FDA in 1990 to provide a unified national focus for collection of all reports of clinically significant adverse events, including but not limited to those mandated for reporting.[142] To increase sensitivity, the VAERS form is designed to permit narrative descriptions of adverse events. All people, including patients or their parents and not just health-care professionals, are permitted to report to VAERS, especially clinically significant events. As of 2010, 15% of US VAERS reports come directly from consumers. There are no restrictions set on the interval between vaccination and onset of illness or requirements that a patient must have medical care in order for the event to be reported. Web-based reporting became available in 2002; experience to date shows it to be more complete and timely[208] and it was therefore heavily used during the 2003 US smallpox[209] and 2009 H1N1 vaccination campaigns.[210] Future potential enhancements include: (i) integration of VAERS reporting modules with computerized immunization registries that can transfer vaccine exposure and patient identifier information automatically, resulting in more accurate, complete, efficient, and timely transmission of VAERS reports;[211] and (ii) reporting of denominators from the registry to allow calculation of VAERS reporting rates.[212] The latter is especially important to overcome the problem of interpreting VAERS data in the face of increasing heterogeneity of vaccine exposures in the United States.

Enhancements to VAERS passive surveillance since its inception have included capability for near real-time report review by CDC and FDA medical officers, collaborations with professional medical associations, development of a user-friendly public use data query tool, and web-based reporting, in

Table 26.2 Table of reportable events following vaccination, United States*

Vaccine/toxoid	Event	Interval from vaccination
Tetanus in any combination; DTaP, DTP, DTP-HiB, DT, Td, TT, Tdap, DTaP-IPV, DTaP-IPV/Hib, DTaPHepB-IPV	A. Anaphylaxis or anaphylactic shock B. Brachial neuritis C. Any sequela (including death) of above events D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	7 days 28 days Not applicable See package insert
Pertussis in any combination; DTaP, DTP, DTP-HiB, Tdap, P, DTaP-IPV, DTaP-IPV/Hib, DTaPHepB-IPV	A. Anaphylaxis or anaphylactic shock B. Encephalopathy (or encephalitis) C. Any sequela (including death) of above events D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	7 days 7 days Not applicable See package insert
Measles, mumps, and rubella in any combination; MMR, MR, M, MMRV, R	A. Anaphylaxis or anaphylactic shock B. Encephalopathy (or encephalitis) C. Any sequela (including death) of above events D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	7 days 15 days Not applicable See package insert
Rubella in any combination; MMR, MMRV, MR, R	A. Chronic arthritis B. Any sequela (including death) of above events C. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	42 days Not applicable See package insert
Measles in any combination; MMR, MMRV, MR, M	A. Thrombocytopenic purpura B. Vaccine-strain measles viral infection in an immunodeficient recipient C. Any sequela (including death) of above events D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	30 days 6 months Not applicable See package insert
Oral polio (OPV)	A. Paralytic polio B. Vaccine-strain polio viral infection C. Any sequela (including death) of above events D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	30 days†/6 months‡ 30 days†/6 months‡ Not applicable See package insert
Inactivated polio (IPV) in any combination; DTaP-IPV, DTaP-IPV/HIB, DTaP-HepB-IPV	A. Anaphylaxis or anaphylactic shock B. Any sequela (including death) of the above event C. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	7 days Not applicable See package insert
Hepatitis B in any combination; with HepB, HepAHepB, DTaP-HepB-IPV, Hib-HepB	A. Anaphylaxis or anaphylactic shock B. Any sequela (including death) of the above event C. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	7 days Not applicable See package insert
Hemophilus influenza type b in any combination; (conjugate)-Hib, Hib-HepB, DTP-Hib, DTaPIPV/Hib	Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert

(*Continued*)

Table 26.2 (*Continued*)

Vaccine/toxoid	Event	Interval from vaccination
Varicella in any combination; VAR, MMRV	Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Rotavirus (monovalent or pentavalent) RV1, RV5	Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Pneumococcal conjugate (7-valent or 13-valent) PCV7, PCV13	Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Hepatitis A in any combination; HepA, HepA-HepB	Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Influenza – trivalent inactivated influenza, live attenuated influenza – TIV, LAIV	Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Meningococcal - MCV4, MPSV4	Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Human papillomavirus (Quadrivalent or Bivalent) - HPV4, HPV2	Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert

* Effective November 10, 2008. The Reportable Events Table (RET) reflects what is reportable by law (42 USC 300aa-25) to the Vaccine Adverse Event Reporting System (VAERS) including conditions found in the manufacturers package insert. In addition, individuals are encouraged to report *any* clinically significant or unexpected events (even if you are not certain the vaccine cause the event) for *any* vaccine, whether or not it is listed on the RET. Manufacturers are also required by regulation (21CFR 600.80) to report to the VAERS program all adverse events made known to them for any vaccine. A list of vaccine abbreviations is located at: http://www.cdc.gov/vaccines/recs/acip/vac-abbrev.htm

† In a non-immunodeficient recipient.

‡ In an immunodeficient recipient.

addition to more frequent review and dissemination of safety data. "Enhanced passive" surveillance via VAERS has been successfully used to date in safety surveillance for rotavirus,[213] yellow fever,[214] smallpox vaccines,[215] and 2009 H1N1 vaccine,[210] and would likely be implemented in any counter-bioterrorism-related, wide-scale vaccination program.[216]

Among the approximately 27,000 US VAERS reports now received annually, about 9% are classified as serious (reported as resulting in death, life-threatening illness, disability, or hospitalization).[217] A contractor, under CDC and FDA supervision, collects, codes (using the Medical Dictionary for Regulatory Activities, MedDRA),[218] and enters VAERS reports in a database. Trained nurses follow up with the person who reported an event classified as serious to obtain additional medical information and recovery status. CDC and FDA have access to the VAERS database and focus their efforts on analytical tasks of interest to the respective agencies. A database of initial VAERS reports (without personal identifiers) and a user-friendly data query tool are available to the public at www.vaers.hhs.gov.

Other national experiences

Several other countries also have substantial experience with passive surveillance for vaccine safety.

What is now the Canadian Adverse Events Following Immunization Surveillance System (CAEFISS) was first developed in 1987.[195] Reporting forms have check-off boxes for specific events with accompanying case definitions. Provision is also made for an "other" category. To supplement the passive system, an active, pediatric hospital-based surveillance system that searches all admissions for possible relationships to immunizations, known as Immunization Monitoring Program–Active (IMPACT), has been operational since 1990.[219,220] An Advisory Committee on Causality Assessment, consisting of a panel of experts, has also been formed to review the serious passive reports.[221] The Netherlands also convenes a panel of experts annually to categorize their reports, which are then published.[222] The United Kingdom and most members of the Commonwealth use the "yellow card" system, where a reporting form is attached to officially issued prescription pads.[223] 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.[192,224]

A field guide for implementation of monitoring of AEFI has been developed by WHO.[225] The primary focus is on detection of correctable programmatic errors such as injection site abscesses (suggestive of inadequate sterilization), and development of a rapid response/ assessment team for clusters of more serious events (e.g., toxic shock syndrome from contamination of vaccine vials[198] or deaths from confusing other medications for vaccines[226]). As more new vaccines are first introduced in low and middle-income countries, there is increasing awareness of the need to improve currently inadequate pharmacovigilance systems in these countries.[11] The decades long delay in discovering serious AEFIs after yellow fever vaccination,[227] and BCG vaccination in human immunodeficiency virus-infected infants[228] further highlight this urgent need.

Classifications and case definitions

Vaccine adverse events can be classified by frequency (common, rare), extent (local, systemic), severity (hospitalization, disability, death), causality (probable, possible, unlikely, etc.),[203] and preventability (intrinsic to vaccine, faulty production, faulty administration). Wilson developed the first classification system with focus on errors of production (e.g., bacterial, viral, toxin contamination) and administration (e.g., non-sterile apparatus).[5] A more recent classification[229,230] divides adverse events after vaccinations into: (i) *vaccine-induced*: due to the intrinsic characteristic of the vaccine preparation and the individual response of the vaccinee, these events would not have occurred without vaccination (e.g., vaccine-associated paralytic poliomyelitis); (ii) *vaccine-potentiated*: may have occurred anyway, but were precipitated by the vaccination (e.g., first febrile seizure in a predisposed child); (iii) *programmatic error*: due to technical errors in vaccine preparation, handling, or administration; or (iv) *coincidental*: associated temporally with vaccination by chance or due to underlying illness. The distinction between vaccine-induced and vaccine-potentiated, as first clarified for DTP and DT vaccine and infantile spasm,[231] has been useful because vaccine-potentiation does not result in excess vaccine-attributable risk over time, whereas vaccine-induced does (Figure 26.1).

The Dutch system further classifies a report based on whether single or multiple vaccines were received and single or multiple adverse events were reported.[232] Case definitions of certain vaccine adverse events were first developed in Brazil,[233] Canada,[234] India,[198] and the Netherlands.[232] When case definitions were added to the Canadian form as guidance for what should be reported, the proportion of reports meeting the case definition criteria increased from 69 to 87%.[141] Alternatively, in a more open reporting system such as VAERS, these definitions can be applied to reports to develop a case series for further investigation.[235,236] Real progress in implementation of similar standards across national boundaries are being realized with the advent of the International Conference on Harmonization (ICH)[237] and the Brighton Collaboration.[143]

The Brighton Collaboration was established in 2000 and is an international voluntary effort to facilitate the development, evaluation, and

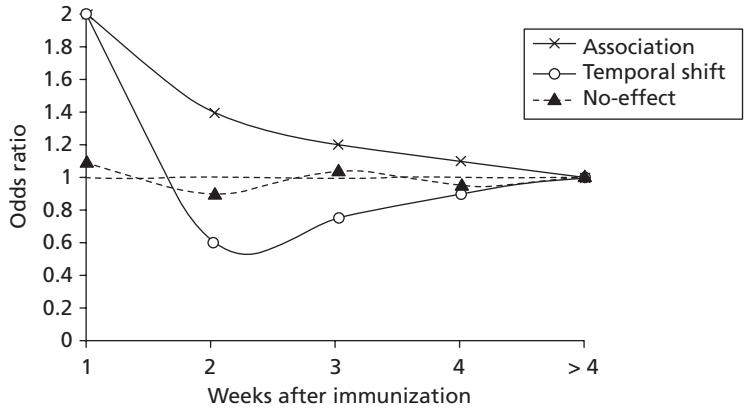


Figure 26.1 Three theoretical models of the temporal relationship between immunization and an adverse effect: (1) Association: the risk exceeds 1 at all time windows postimmunization; (2) temporal shift: the risk exceeds 1 initially but then falls below 1 but coming back to 1 eventually, such that the area under the curve above and below 1 is similar; and (3) no effect: the risk stays around 1.[231]

dissemination of standardized case definitions of Adverse Events Following Immunizations.[143] Global workgroups of experts are convened to develop case definitions that are then reviewed by relevant experts. The Brighton case definitions for each adverse event 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 developing to developed country settings). Over 30 Brighton case definitions are now available for use at www.brightoncollaboration.org.

Standardized clinical assessment protocols and centers

There has been an increasing awareness that the utility of a spontaneous reporting system (SRS) as a potential disease registry and the immunization safety infrastructure can be usefully augmented by tertiary clinical centers. The United States initiated its military Vaccine Healthcare Centers (VHC; www.vhcinfo.org) and civilian Clinical Immunization Safety Assessment (CISA) Network in 2001; these bring together infectious disease epidemiologists, immunologists, dermatologists, and other subspecialists from multiple participating sites as needed for various tasks.[32] Among these tasks is the standardized assessment of people who suffered an apparently true vaccine reaction (e.g., hypersensitivity),[48,238] to improve our scientific understanding of the pathophysiology and risk factors of the reaction. New understanding of the human genome, pharmacogenomics, and immunology may now make it possible for us to truly understand these reactions (see also Chapter 34).[129,239] For example, studies of myopericarditis following smallpox vaccination in the VHC suggested increased risk in people with HLA type UD,[19] and a person with a case of yellow fever vaccine-associated viscerotropic disease showed polymorphisms in *CCR5* and *RANTES* genes.[240]

Through these centers, standardized assessment protocols can be developed to examine patients with similar adverse events 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. The diagnosis of a specific epilepsy syndrome was made in 14 cases with alleged pertussis vaccine encephalopathy; genetic mutations of the sodium channel gene were identified in 11 of 14 patients.[61]

For patients who have had adverse events that generate concern but do not contraindicate com-

pletion of a vaccine series, such as hypotonic–hyporesponsive episodes[129] and extensive limb swelling after acellular pertussis vaccination,[133,241] the standardized clinical assessment centers, such as CISA, can provide assessment and management of subsequent vaccinations under protocols.

Finally, standardized clinical assessment centers can provide regional referral and advice services—with the opportunity to follow-up and document compliance with advice provided and outcome so that this 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).[242] Both 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: (i) *Can It?*; (ii) *Did It?*; and (iii) *Will It?*.[243] The answer to *Can It?* was the focus of the Institute of Medicine reviews.[27,28,40a] 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.[244]

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 adverse event 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").[28] Information useful for assessing causality in individual case reports include: (i) previous general experience with vaccine (e.g., duration of licensure, number of vaccinees, whether similar events have been observed among other vaccinees or non-vaccinees, existence of animal models to test vaccine as a cause); (ii) alternative etiologies; (iii) biologic plausibility; (iv) individual characteristics of the vaccinee that may increase the risk of the adverse event; (v) timing of events; (vi) characteristics of the event (e.g., laboratory findings); and (vii) rechallenge[245,246] (see also Chapter 33).

When a vaccine *can* cause an adverse event, the *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 contraindications for the individuals and risk–benefit policy decisions for the population. The *Will It?* is usually very difficult to answer, however, as it can only be answered based on epidemiologic studies.[28] Furthermore, the sample sizes of such studies may be large enough to establish whether vaccine can cause a given event but yet inadequate to stratify by subgroups to examine risk factors that can help delineate potential contraindications.

Specific adverse events may be considered to be caused by a specific vaccine if the event is associated with (i) a unique laboratory finding, and/or (ii) 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 vaccine and not wild strain by genetic sequencing.[144] The detection of IgG antibodies to the stabilizers in vaccine in children with hypersensitivity reactions confirms the etiology.[46,49] 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.[247] Acute flaccid paralysis, especially shortly after receipt (or contact with a recipient) of oral polio vaccine, is almost pathognomonic of vaccine-associated paralytic polio in countries where wild polio virus is unlikely to be

circulating.[56,248] Similarly, acute myopericarditis in otherwise healthy recent smallpox vaccinees also supports a causal relationship.[18,19] 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.[138] But unlike 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 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.[27] This relationship is not universal, however, as pregnant women who receive live attenuated rubella vaccine, unlike those exposed to wild rubella, have not been shown to have illness compatible with congenital rubella syndrome.[249] Clustering of events in time after vaccination can also suggest causation if "reporting bias" can be ruled out. Such bias may occur as parents and doctors are most likely to link adverse events 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.[103,250] Onset of GBS after the swine influenza vaccination was delayed up to 6 weeks, but clustered at 2 to 3 weeks following vaccination, as autoimmune demyelination is a slower process.[54] 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 26.1).[231,251]

Unfortunately, most serious reported vaccine adverse events lack these unique features that permit easy inferences on causality. Adverse events such as 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 simplify their administrative proceedings by making a blanket assumption that all adverse events 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.[92] Despite these caveats, the timing of the onset interval after vaccination plays a major role in most causality assessment algorithms, as AEFIs after live viral vaccines usually occur later than those of killed vaccines. The WHO classifies a clinical event with a plausible time relationship to vaccine administration and which cannot be explained by concurrent disease or other drugs or chemicals as *Very Likely/ Certain* to be caused by vaccination; a clinical event with a reasonable time relationship to vaccine administration and is unlikely to be attributed to concurrent disease or other drugs or chemicals as *Probable*; and that which could also be explained by concurrent disease or other drugs or chemicals as *Possible*.[203,252]

In some countries, expert committees of specialists in relevant disciplines (e.g., pediatrics, infectious disease, neurology) review reports. This "global introspection" approach[253] has been used in both Canada[221] and the Netherlands[222] to classify reports of adverse events in gradations of probable association to vaccination (see also Chapter 33). The CISA network used a standardized protocol for individual case reviews of reported H1N1 vaccine adverse events,[222a] building on the lessons of the Canadian Advisory Committee on Causality Assessment (ACCA). Classifications are based on the reported symptoms, the interval between vaccination and onset of symptoms, and a set of case definitions. Because 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 and Communication (VAESCO) project (www.vaesco.net) recently 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."

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.[254] As indicated above, unusual clinical features and/or clustering in time or space usually suggest that something may be awry. No illness other than GBS was reported more commonly in the second and third week than in the first week after swine influenza vaccination, leading to further validation studies.[54,255,256] Traditionally, a signal occurs when an observed number of events exceeds the number of events expected by chance alone for the specific data source. In general, an acceptable false positive 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.[257,258]

Several recent examples in the United States and elsewhere highlight the importance of rapidly identifying and responding to serious AEFIs identified following new vaccines or newly reintroduced vaccines. After a prelicensure signal,[259] passive reports to VAERS of intussusception among children vaccinated with rhesus rotavirus vaccine were the first postlicensure signal of a problem,[213] leading to several studies to verify these findings.[66,67] Similarly, initial reports to VAERS of a previously unrecognized serious yellow fever vaccine-associated viscerotropic disease,[260,261] and neurotropic disease[214] have since been confirmed elsewhere[227] and as early as 1973, based upon retrospective review.[262] Acute myopericarditis has been a relatively unexpected finding among people vaccinated against smallpox in the United States for bioterrorism preparedness.[19,215] Oculorespiratory syndrome was found among influenza vaccines from one Canadian manufacturer in one season.[132] Bell's palsy was detected in recipients of a new Swiss intranasal influenza vaccine.[106] While several GBS cases were reported to VAERS after introduction of MCV4 in adolescents in the United States,[263] subsequent large controlled studies found no association.[264] Most recently, febrile seizures in young children were observed more than expected in passive reporting in Australia following administration of one formulation of trivalent inactivated influenza vaccine (TIV).[265] A febrile seizure signal was also identified in VAERS after receipt of a different US TIV; after evaluation in the VSD project, the signal was verified and it found that simultaneous administration of another vaccine with TIV contributed to the increased risk.[265a]

Because of the success in detecting these signals, there have been various attempts to automate screening for signals using SRS reports. Historically, this has been relatively unsuccessful,[266] largely because of inherent methodologic problems of spontaneous reports (see above 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 are increasingly being applied.[267–270] VAERS is one of the largest registries for rare vaccine adverse events in the world, with approximately 300 000 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 made to apply various computer-assisted techniques for automated detection of unusual trends and patterns. Several different "data mining" methods that have been evaluated in VAERS to date, include Empirical Bayesian,[271] Association Rule Discovery,[272] multi-item gamma Poisson shrinkage (MGPS),[273] proportional morbidity distribution,[274] and proportional reporting rate ratio.[268,275] No single method appears to be superior.[276] Rational approaches to prioritizing the large numbers of potential signals generated using automated algorithms on large passive AEFI report databases may involve utilization of complementary approaches, such as data visualization and an array of different data mining methods (each with pros and cons), where cumulative higher score might signal cause for greater concern. Ultimately, these methods do represent a useful adjunct to, but not a substitute for, traditional methods of scrutinizing spontaneous reports in increasingly complex databases such as VAERS.[268]

A "rapid cycle" analytic approach has been developed by the CDC VSD project (see Automated Large, Linked Databases, below) to rapidly analyze safety data on new vaccines,[277] existing vaccines with new recommendations or indications, and the annual influenza vaccine strain.[278] This initiative utilizes the strengths of the VSD with its ability to gather automated vaccination and medical care utilization data from enrolled members in ten managed care organizations, and incorporates new data management to collect and analyze the safety profile of each successive week's cohort of vaccinated people. To date, the project has not only successfully simulated,[278–280] but also detected an observed increase in febrile seizures after combination MMRV vaccines.[281] The VSD covers a population of about 9 million people, which while large, may still not be able to detect associations between rare exposures and very rare outcomes, or more common exposures or outcomes among a specific subpopulation, such as pregnant women. Furthermore, rapid cycle analysis requires *a priori* specification of the adverse events to be studied, and therefore is not a true source of *de novo* signals. New information theory approaches may provide a way of detecting previously unexpected associations after vaccination.[282] Until then, 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 result either in more prominent clusters of vaccine adverse events, 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.[283] Surveillance of vaccine adverse events 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,[49] viscerotropic disease following yellow fever vaccine,[284] aseptic meningitis after mumps vaccine,[104,285] GBS with swine influenza vaccine,[54] GBS after oral polio vaccine,[286] allergic reactions after Japanese encephalitis vaccine,[287] neuropathy after rubella vaccine[288]) or absent (e.g., events after meningococcal vaccine,[289,290] GBS after measles[291]). Such signals still require validation, however, since some, after more careful scientific studies, are not confirmed to represent a true association.[292,293] Mass psychogenic illness can plague mass vaccination campaigns, especially among adolescents in school settings.[294,295]

Preparation in advance of mass vaccination campaigns is critical. During mass campaigns with group B meningococcal vaccine in New Zealand[296] and during the large vaccination effort for 2009 H1N1 influenza vaccine in the United States and elsewhere,[55,297,298] several systems were put in place to identify signals early. For the latter, in the United States, 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 H1N1vaccination. Both systems had strength in the population size and the rapid review of reports.[55] Assessing and having background rates for likely AEFIs during mass campaigns is also very helpful.[298,299] Special registries or studies will be needed, however, to monitor the outcome for subpopulations such as pregnant women, who may need to be vaccinated with limited safety data during such campaigns.[300]

Lessons learned to date

Several lessons are beginning to emerge from spontaneous reporting systems such as VAERS.[217,301–303] Such systems worldwide have successfully detected previously unrecognized reactions and helped to obtain data to evaluate whether AEFIs are causally linked to vaccines.[106,132,138,213–215] VAERS has also successfully served as a source of cases for further investigations of idiopathic thrombocytopenic purpura after MMR,[304] anaphylaxis after MMR,[46] and syncope after immunization.[305] 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.[306,307]

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, reported 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. 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 such as hospitalization and seizures after DTaP in toddlers was one-third that after DTP.[274] Reports to VAERS of vaccine-associated paralytic polio disappeared after the shift away from oral polio vaccine in the United States.[217] The proportion of GBS reports following inactivated influenza vaccines over several seasons did not vary including following 2009 H1N1 vaccines even though the reporting rates for GBS were higher following 2009 H1N1 which was likely attributed to heightened media attention.[217a] 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 adverse events, including whether certain lots of vaccines are associated with unusually high rates of adverse events,[308,309] especially when combined with estimates of lot use denominator obtained from statistical models.[310]

Data from passive spontaneous reporting systems, such as VAERS, have helped to inform potential clinical management[311] of vaccine adverse events and to identify potential risk factors for such events, such as advanced age[261] and thymic dysfunction[312] associated with yellow fever vaccine complications, concurrent zoster infection in varicella vaccinees resulting in meningitis,[313] personal and family history of convulsions in pertussis vaccinees,[108] and factors associated with postvaccinial syncope-related injuries.[305,314]

The reporting efficiency or sensitivity of a spontaneous reporting system can be estimated if the expected rates of adverse events generated from carefully executed studies are available. A study using this method showed a higher proportion of serious events such as seizures that follow vaccinations are likely to be reported to VAERS (or its predecessor, Monitoring System for Adverse Events Following Immunizations, MSAEFI) than milder events such as rash, or delayed events requiring laboratory assessment, such as thrombocytopenic purpura after measles–mumps–rubella vaccination (Table 26.3).[315] "Capture–recapture" methods, when at least two independent sources are available to ascertain incident adverse event cases in the same population and enough identifying data on the cases are also available to identify individuals ascertained in both datasets 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.[316] Although formal evaluation has been limited, the probability that a serious event reported to VAERS 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%).[317] In general, the validity of diagnoses reported to VAERS is highly variable, depending on the condition.

Despite the above uses, spontaneous reporting systems for drug and vaccine safety have a number of major methodologic weaknesses (see also Chapter 10) and pitfalls for the unwary use of public-use datasets.[303] Biased and incomplete reporting are inherent to all such spontaneous reporting systems and potential safety concerns may be missed.[315,318] Aseptic meningitis associated with the Urabe mumps vaccine strain, for example, was not detected by spontaneous reporting systems in most countries until eight cases with vaccine-specific virus isolation were published in 1989, 7

Table 26.3 Reporting efficiencies for selected outcomes, two passive surveillance systems for vaccine adverse events, United States[311]

Adverse event	Vaccine	Reporting efficiency(%)		
		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

*MSAEFI, Monitoring System for Adverse Events Following Immunizations; VAERS, Vaccine Adverse Event Reporting System.
†Public and private sector information is missing on these cases.

Box 26.1 "2 × 2" table necessary for epidemiological analysis of causality between vaccine and an adverse event

		Adverse event	
		Yes	No
Vaccinated	**Yes**	"a"	"b"
	No	"c"	"d"

- Rate of adverse event following vaccination = a/a+b
- Rate of adverse event in the absence of vaccination = c/c+d
- Reports to passive surveillance systems for vaccine adverse events (e.g., Vaccine Adverse Event Reporting System) represent just partial information (due to under- and biased reporting) for cell "a" of the table. Epidemiologic studies aim to gather information for all four cells of this table in an unbiased manner.

years after licensure.[319] Most importantly, however, the information content of such spontaneous reports represent just cell "a" of a two-by-two table of vaccination versus adverse event (Box 26.1), and an incomplete and biased content at that.[32]

Use of data from spontaneous reporting systems 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. Since much of "signal detection" relies on specific diagnoses and their coding into databases, new adverse event clinical syndromes may not be "recognized" and analyzed as such until hypothesis strengthening procedures such as 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 (ORS) after 2000–2001 influenza vaccination;[132,320] which, in retrospect, probably also occurred in other influenza seasons[321] and other countries with other influenza vaccine manufacturers.[322]

Current spontaneous reporting systems are also prone to detecting increases in adverse event rates that are not true increases. Instead, they may be due to an increase in: (i) reporting efficiency, (ii) vaccine coverage, or (iii) increases in the incidence of known or unknown etiologies for a particular adverse event. Spontaneous reporting systems are usually unable to sort out causally related from coincidentally related adverse events 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.[150] An increased reporting rate of an adverse event 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).[323] A signal of venous thromboembolic events in human papillomavirus vaccinees in VAERS was probably due to confounding from concurrent use of oral contraceptives.[324] Finally, an approximately two- to threefold increase in 2009 H1N1 reports to VAERS as compared to 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 H1N1 vaccine.[325]

These studies 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 increased number of reports has been dropped.[326]

Historically, most (especially resource-limited) countries have relied on spontaneous reporting systems alone for postlicensure vaccine safety monitoring. The inadequacy of scientific information on vaccine safety found by the Institute of Medicine is related to the methodologic weaknesses inherent to spontaneous reporting systems. 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 adverse event rates.[327,328]

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.[329] Alternatively, active surveillance in a large trial can be done to show the attributable risk for a specific adverse event (e.g., intussusception) was lower for a new rotavirus vaccine compared to the old.[115] When new adverse events such as myopericarditis are detected after smallpox vaccination, trials of new vaccine candidates using a similar viral vector may require more safety assessment (e.g., electrocardiogram).[330]

Postlicensure clinical trials

To optimize vaccine use, clinical trials may be conducted after vaccine licensure to assess the effects of changes in vaccine formulation,[331] vaccine strain,[117,332] age at vaccination,[333] the number and timing of vaccine doses,[334] simultaneous administration,[335] and interchangeability of vaccines from different manufacturers[336] 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 developing countries,[337] albeit lower than among unvaccinated children.[338] This finding resulted in a change in recommendations by WHO for the use of such vaccines.[339] The development of automated large, linked databases (see below) may permit 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 adverse events 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,[340–342] DTaP,[118] and varicella vaccines (including multiyear evaluation for disease incidence, herpes zoster, and a pregnancy registry).[343,344] Postapproval studies are now routine for newly licensed vaccines such as MMRV vaccine,[107] human papillomavirus

vaccine,[345] and second-generation rotavirus vaccines.[346,347] 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.[348] Requirements for postapproval evaluation have even been extended to less frequently used vaccines, such as Japanese encephalitis vaccine.[349] A large postlicensure randomized trial for this vaccine was also completed in China to improve the available data on its short-term safety.[350]

Ad hoc epidemiologic studies

Historically, *ad hoc* epidemiologic studies have been employed to assess signals of potential adverse events generated by spontaneous reporting systems, 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 adverse event following vaccination with risk in the absence of vaccination. Occasionally, data collected for other study outcomes may be reanalyzed to see if vaccine was causally related or not.[351] Examples of *ad hoc* follow-up studies to signals of vaccine safety issues are: the investigations of poliomyelitis after inactivated[77] and oral polio vaccines;[248] SIDS after DTP vaccination;[27,153,352–354] encephalopathy after DTP vaccination;[59,355] meningoencephalitis after mumps vaccination;[144,356] injection site abscesses postvaccination;[357] intussusception after Rotashield vaccine;[66,67,66,67] vaccinations and autism;[358,359] GBS after influenza vaccine;[54,55,150] and GBS after meningococcal conjugate vaccine.[264] Many such studies have been compiled and reviewed by the Institute of Medicine.[27,28,34,36–40] While automated large, linked databases (see below) provide a more cost-effective and flexible framework for hypothesis testing, *ad hoc* epidemiologic studies may still be needed in settings without automated large, linked databases,[106,132] or where the statistical power of the automated large, linked databases may be inadequate to answer a question in a timely manner.[149,150,317]

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 type of event. As with drug safety research (see Chapters 11–21), efforts have increasingly turned to record linkage between automated exposure (immunization records in lieu of pharmacy) files and outcome medical files. The CDC participated during the late 1980s in two pilot vaccine safety studies using automated large, linked databases in Medicaid and Managed Care Organizations (MCO) populations, respectively.[360,361] 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.[27] These limitations, the constraints of VAERS, and the recognition of the need for improved monitoring of vaccine safety, prompted the CDC to initiate the Vaccine Safety Datalink (VSD) project in 1990.[70,250,362] 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 in 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 implementation of diagnosis-related group (DRG) billing.[363] To increase patient confidentiality, the VSD shifted from annual data tape submissions from the MCOs for data pooling and analysis at CDC to a distributed network data management model; in parallel, VSD is also increasing transparency via public access data sharing and external input.[364]

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 US population in these age groups).[250] Expansion to eight MCOs (including data on all age groups at three MCOs) was accomplished in 2000.[362] The VSD focused its initial efforts on examining potential associations between immunizations and 34 serious neurologic, allergic, hematologic, infectious, inflammatory, and meta-

bolic conditions. The VSD is also being used to test new *ad hoc* vaccine safety hypotheses that arise from the medical literature,[16,147,148,170,365–367] from VAERS,[67,323] from changes in immunization schedules,[368,369] or introduction of new vaccines.[281,340,342,370] In addition, the VSD databases have been used to conduct influenza vaccine safety studies in which large cohorts of children are screened for evidence of increased medically attended events following vaccination.[167,168] 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.[281,369] At the time of this writing (2010), ongoing surveillance is currently being conducted on the following combination vaccinations: MMR-V, Tdap, DTaP-IPV, and DTaP-IPV/Hib.

When the VSD identifies an adverse event as being associated with vaccine, data on the incidence rate attributable to vaccine are available,[67,170,281] permitting accurate risk–benefit assessment by both the public and policymakers.[371] Subgroup analyses may permit identification of risk factors for adverse events (or vaccine failures), which may be useful in identifying contraindications to vaccinations.[372] 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 SRS.[299] Also, incidence rates of vaccine-associated adverse events derived from VSD can be used to evaluate the sensitivity of passive reporting systems. The VSD data also aid the Vaccine Injury Compensation Program in determinations of what events should be compensated as vaccine "injuries."[92]

In addition to *ad hoc* epidemiologic studies, a Rapid Cycle Analysis (RCA) team was formed within the VSD to conduct near-real-time active surveillance on newly licensed vaccines. The RCA relies on analytic data sets that are created weekly from the automated MCO data. The weekly analytic data sets are used to investigate potential associations between vaccines and adverse events 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.[373] The first application of MaxSPRT was in safety assessment of the newly licensed meningococcal conjugate vaccine in 2005.[277] 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.[281] This finding precipitated changes in US immunization policy for MMRV and MMR + V in children.

The VSD has some limitations. While diverse, the population in the MCOs currently in the VSD is not wholly representative of the United States in terms of geography or socioeconomic status. More importantly, because of the high coverage attained in the MCOs for most vaccines, few non-vaccinated controls are available. Therefore, the VSD often relies on some type of "risk-interval" analysis[360,361,374] (Box 26.2). The capability of this approach to assess associations between vaccination and adverse events with delayed or insidious onset (e.g., neurodevelopmental or behavioral outcomes) is limited.[16] The VSD also cannot easily assess adverse events that do not result in a health-care visit, and therefore are not currently captured in existing MCO databases, because they do not result in a health-care consultation (e.g., fever).[250] 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 epidemiologic methods, it may not successfully control for confounding and bias in each analysis,[121] and inferences on causality may be limited.[375]

Despite these potential shortcomings, the VSD provides an essential, powerful, and relatively cost-effective complement to ongoing evaluations of vaccine safety in the United States. In view of the methodologic and logistical advantages offered by automated, large, linked databases, Denmark,[33] the United Kingdom,[103,144,376,377] and Canada[378] have also developed large automated databases linking immunization registries with medical files. Europe[30] and Taiwan[298] anticipate they will eventually convert their 2009 H1N1 vaccine safety

Box 26.2 Example of method for risk-interval analysis of association between a universally recommended three-dose vaccine (with few unvaccinated people for comparison) and adverse event

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 non-risk interval time = 365 − 90 = 275 person-days).

o———x====———x====———x====———//———>
Birth Dose 1 Dose 2 Dose 3 365 days

3 Add up (i) total risk interval and non-risk 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 below uses 3 and 10 cases):

	Adverse event	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$_{unvaccinated}$ = 0/1000 = 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 for developing the adverse event within the interval following vaccination

surveillance into large, linked databases for routine vaccinations. The first such pilot database in a less developed setting has been established in Vietnam.[379] Given that many vaccines against many poverty related diseases such as rotavirus, malaria, and TB will be introduced first in such countries, there is a need to develop VSD-like infrastructures there, too.[126]

Methodologic approaches for observational epidemiologic 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 United States, 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.[207] This requirement, along with improvements in technology, has prompted many organizations to automate their vaccination records.[328]

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.[380,381] When people receive their vaccinations from several providers (not uncommon in the United States), their exposure status may be misclassified.[382] This error could be minimized if a centralized Immunization Information Systems (IIS) were implemented to track all vaccinations from birth. Such an IIS has been implemented in Denmark,[33] most of the

United Kingdom,[327] and state/regional IIS are under development in the United States,[328] Canada,[383] and Australia.[384]

The availability and quality of vaccination records generally decrease as people age. Some vaccines for older people (e.g., tetanus–diphtheria boosters in emergency rooms, hepatitis B vaccinations for health-care personnel) may be administered in settings other than primary health care. 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 adverse event 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,[385] medical records from the village, the hospital, and the regional public health nurse, in addition to the automated vaccination record, were reviewed.[385,386] Studies of GBS and H1N1 influenza vaccine relied on patient/family interview, hospital medical record, and/or validation with primary care providers for exposure ascertainment.[55] Interestingly, reliance on provider verification may lead to under-ascertainment of vaccination status, either because of poor record keeping[382] or concerns about liability in vaccine safety studies.[150]

Standards should be developed to further improve the accuracy and efficiency of transfer of vaccine identification information from the vaccine vial to either automated or paper immunization records, including: (i) abbreviations for new vaccine antigens and vaccine manufacturers, (ii) peel-off labels, (iii) bar codes, (iv) lot numbers, (v) immunization records, and (vi) presentation of key identifier information on vaccine packaging (as on the nutrition label). WHO has recently identified as a priority the development of a vaccine dictionary that will allow differentiation of vaccine formulations from various manufacturers.[387]

Outcomes

To ensure both high sensitivity and specificity for an AEFI, a multistep approach is usually required for case ascertainment.[167,168,170,365,388] In step one, the automated databases are screened to identify ICD-9 diagnostic codes for the condition of interest. The ICD-9 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, the medical records of the patients are often reviewed by a trained abstractor, blinded to vaccination status. 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 such as 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.[161,389] 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 such as GBS, a panel of specialists may also be asked to review the medical records after exposure status has been masked.[65]

This process minimizes the likelihood of a false negative conclusion (due to bias towards 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[213] may have been just the tip of the "iceberg" of a broader syndrome that also included bloody stool, vomiting, diarrhea, and abdominal pain.[306] Adverse neurologic outcomes other than GBS were reported among the 1976–1977 and 2009 H1N1 influenza vaccinees.[390,391] 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.[392]

In the context of real-time surveillance, influenza vaccine safety monitoring is hindered by the rate at which large, linked databases 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 4 months.[393] For influenza vaccine safety monitoring, the influenza season may be over by the time the outcome data are fully captured, thereby rendering the real-time analysis moot.

Study design and analytic methods

Different analytical 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 non-vaccinees (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, MCO, age, high-risk condition, and calendar time. The cohort of vaccinated and unvaccinated individuals is 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 current understanding of the most plausible biologic mechanism, should such an association actually exist. For most acute events, exposure windows of 0–2, 1–14, 1–30, and 1–42 days are often used.[157,166,168,170] 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.[394,395] 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 large, linked databases such as ethnicity, socioeconomic status, and underlying health state.[121]

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, MCO site, and calendar time.[66,160,396] 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.[394]

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.[66,106,166,170,173,174,397–401] 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 non-risk time periods are compared, but only vaccinated individuals are included in the analysis. A time period immediately following vaccination is defined as the 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 non-risk (or control) 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 non-risk time periods are compared, but only cases with an event are included in the analysis.[173,358,397,398,402] 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 below, the SCCS has nearly as much statistical power as the cohort approach when a high proportion of the population is vaccinated.

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.[397,400] Mis-specifying such variables can lead to biased results—particularly when the event is rare.[402] Alternatively, it has also been shown that semiparametric 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.[398]

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. A recent analytic method has been developed to account for the postevent dependence in an SCCS analyses when the postvaccination risk period is short and when the event is both rare and non-recurrent.[399] 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.[174,397] In a study using VSD data and simulated cases of a rare, acute illness (immune thrombocytopenic purpura or ITP) after MMR vaccination, 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 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.[280] 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 a 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 economic 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.[106,170,393,400] 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.[393,403] 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 dilemma with any design, the SCCS and risk-interval designs are particularly susceptible to this type of seasonal bias.

In addition to seasonality, studying the safety of influenza vaccination is challenged by the potential for selection bias, since it can be assumed that individuals who receive influenza vaccination are different from those who do not. For example, for those vaccinated early, a large percentage of the vaccinated may have co-morbidities, placing them at high risk for infection resulting in them being targeted for vaccination; for those vaccinated later, after universal recommendations, the inherent differences between the vaccinated and unvaccinated populations may change. Moreover, in large, linked MCO databases, it is possible that a certain proportion of the population received an influenza vaccination outside of the MCO, which may not be captured in the automated databases.[393] As described earlier, these potentials for selection bias and exposure misclassification are problematic for the cohort and case–control designs.

The future

Although considerable progress has been made in the development of vaccine safety analytic methods, several challenges remain. Areas of particular importance include the following: (i) minimizing seasonal time-varying bias, (ii) identifying optimal risk windows, (iii) detecting a lifetime dose response from multiple influenza and tetanus-containing vaccinations, (iv) evaluating the safety of simultaneous vaccination, and (v) data mining for unknown AEFIs in real-time active surveillance.

As described, accounting for seasonal time-varying bias is particularly challenging in safety studies of influenza vaccination. Because a large majority of influenza vaccinations are administered in October and November, there may not be enough variability in the temporal distribution of vaccination to control for seasonal fluctuations in the outcome of interest. A potential strategy is the case-centered approach.[403] This strategy has been used to assess influenza vaccine effectiveness in the elderly. In cases occurring during an influenza season, the method uses data from the entire cohort (cases and non-cases) 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.

Future work needs to focus on how to apply this method in studies of vaccine safety, where the study population is young and healthy, and both the exposure and outcome are acute and transient. Moreover, the method could be developed for use in real-time active surveillance of newly licensed vaccines.

Although risk window lengths are often based on prior biologic knowledge, they are also somewhat arbitrarily defined (e.g., 0–2, 1–14, 1–42 days after vaccination). Inaccurate specification of the risk window can result in either 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 versus 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 has shown that there is a linear relationship between the calculated risk and risk window length.[404] The analytic approach calculates an optimal risk window length based on maximum likelihood methods and the study design of interest. Thus far, this approach has been applied to the SCCS design with conditional Poisson regression. 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 adverse events 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 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 adverse events, new methods for disentangling the correlation between dose and age are needed.

An increasing number of parents are choosing to either decline or delay immunizations for their children.[405,406] This implies that there may be a certain amount of variability in the timing at which routine childhood vaccines are administered in the first 5 years of life. This variability may in turn present a natural experiment where the risk of adverse events in an otherwise healthy population of children on alternative schedules can be compared to a cohort of healthy children on the recommended schedule. It is also possible, however, that children on alternative schedules have different health-care utilization patterns than children who are up-to-date, thereby creating a selection bias.[407] Large, linked databases represent an ideal resource to explore this potential natural experiment.

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.[168,408] 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 large, linked databases with real-time active surveillance will address some of these limitations. Such methods would be a natural complement to targeted active surveillance, in which adverse events 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 rates) 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 large, linked databases to identify unsuspected adverse events in real time are needed. These methods should be sensitive enough to detect potentially serious adverse events 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 to be in place should a signal arise.[409]

Acknowledgement and disclaimer

The authors wish to thank Dr David McClure for his contribution to section on Study Design and Analytic Methods, Drs Frank Destefano and Deb Gust for their review, and Dr James Baggs for his assistance with the reference software. 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.

References

1. Brenzel L, Wolfson L, Fox-Rushby LJ, Miller MA, Halsey NA. Vaccine-preventable diseases. In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, *et al.*, eds. *Disease Control Priorities in Developing Countries*, 2nd edn. Washington D.C.: Oxford University Press, 2006, p. 389–412.
2. Miller MA, Hinman AR. Economic analyses of vaccine policies. In: Plotkin S, Orenstein WA, editors. *Vaccines,* 5th edn. Philadelphia: Saunders, 2008, p. 1593–609.
3. Centers for Disease Control and Prevention. Achievements in public health, 1900–1999 impact of vaccines universally recommended for children—United States, 1990–1998 . *MMWR Morb Mortal Wkly Rep* 1999; **48**: 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, Bilous J, Birmingham M, Kieny MP, *et al.* Immunization safety priority project at the World Health Organization. *Semin Pediatr Infect Dis* 2003; **14**: 233–9.
9. Cooper LZ, Larson HJ, Katz SL. Protecting public trust in immunization. *Pediatrics* 2008; **122**: 149–53.
10. Freed GL, Clark SJ, Butchart AT, Singer DC, Davis MM. Parental vaccine safety concerns in 2009. *Pediatrics* 2010; **125**: 654–9.
11. Global Advisory Committee on Vaccine Safety (GACVS), WHO secretariat. Global safety of vaccines: strengthening systems for monitoring, management and the role of GACVS. *Expert Rev Vaccines* 2009; **8**: 705–16.
12. Gangarosa EJ, Galazka AM, Wolfe CR, Phillips LM, Gangarosa RE, Miller E, *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**: 517.
14. Institute of Medicine. Liability for the production and sale of vaccine. In: Sanford JP, ed. *Vaccine Supply and Innovation*. Washington, DC: National Academy Press, 1985, pp. 85–122.
15. Ball LK, Ball R, Pratt RD. An assessment of thimerosal use in childhood vaccines. *Pediatrics* 2001; **107**: 1147–54.
16. Verstraeten T, Davis RL, DeStefano F, Lieu TA, Rhodes PH, Black SB, *et al.* Safety of thimerosal-containing vaccines: a two-phased study of computerized health maintenance organization databases. *Pediatrics* 2003; **112**: 1039–48.
17. Committee to Assess the Safety and Efficacy of the Anthrax Vaccine. *The Anthrax Vaccine: Is it safe? Does it work?* Washington, DC: National Academy Press, 2002.
18. Chen RT, Lane JM. Myocarditis: the unexpected return of smallpox vaccine adverse events. *Lancet* 2003; **362** (9393): 1345–6.
19. Halsell JS, Riddle JR, Atwood JE, Gardner P, Shope R, Poland GA, *et al.* Myopericarditis following smallpox vaccination among vaccinia-naive US military personnel. *JAMA* 2003; **289**: 3283–9.
20. Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson DM, Malik M, *et al.* Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 1998; **351** (9103): 637–41.
21. Smith A, Yarwood J, Salisbury DM. Tracking mothers' attitudes to MMR immunisation 1996–2006. *Vaccine* 2007; **25**: 3996–4002.
22. Balinska MA. Hepatitis B vaccination and French Society ten years after the suspension of the vacci-

nation campaign: how should we raise infant immunization coverage rates? *J Clin Virol* 2009; **46**: 202–5.
23. Clements CJ, Greenough P, Shull D. How vaccine safety can become political—the example of polio in Nigeria. *Curr Drug Saf* 2006; **1**: 117–9.
24. Jegede AS. What led to the Nigerian boycott of the polio vaccination campaign? *PLoS Med* 2007; **4**: e73.
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**: 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**: 33–6.
27. *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 Academy Press, 1991.
28. *Adverse Events Associated with Childhood Vaccines: Evidence Bearing on Causality*. Washington, DC: National Academy Press, 1994.
29. Offit PA, Davis RL, Gust D. Vaccine safety. In: Plotkin S, Orenstein WA, Offit PA, eds. *Vaccines*, 5th edn. Philadelphia: Saunders, 2008, pp. 1629–50.
30. 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**: 371–80.
31. Miller E, Waight P, Farrington P. Safety assessment post-licensure. *Dev Biol Stand* 1998; **95**: 235–43.
32. Chen RT. Evaluation of vaccine safety after the events of 11 September 2001: role of cohort and case-control studies. *Vaccine* 2004; **22**: 2047–53.
33. Hviid A. Postlicensure epidemiology of childhood vaccination: the Danish experience. *Expert Rev Vaccines* 2006; **5**: 641–9.
34. Immunization Safety Review Committee. *Immunization Safety Review: Thimerosal-Containing Vaccines and Neurodevelopmental Disorders*. Washington, DC: Institute of Medicine, National Academy Press, 2001.
35. Immunization Safety Review Committee. *Immunization Safety Review: Measles-Mumps-Rubella Vaccine and Autism*. Washington, DC: Institute of Medicine, National Academy Press, 2001.
36. Immunization Safety Review Committee. *Immunization Safety Review: Multiple Immunizations and Immune Dysfunction*. Washington, DC: Institute of Medicine, National Academy Press, 2002.
37. Immunization Safety Review Committee. *Immunization Safety Review: SV40 Contamination of Polio Vaccine and Cancer*. Washington, DC: Institute of Medicine, National Academy Press, 2002.
38. Immunization Safety Review Committee. *Immunization Safety Review: Hepatitis B Vaccine and Demyelinating Neurological Disorders*. Washington, DC: Institute of Medicine, National Academy Press, 2002.
39. Immunization Safety Review Committee. *Immunization Safety Review: Vaccinations and sudden unexpected death in infancy*. Washington, DC: Institute of Medicine, National Academy Press, 2003.
40. Immunization Safety Review Committee. *Immunization Safety Review: Influenza Vaccines and Neurological Complications*. Washington, DC: Institute of Medicine, National Academy Press, 2003.

40a. Immunization Safety Review Committee. *Adverse Effects of Vaccines: Evidence and Causality*. Washington DC Institute of Medicine, National Academy Press, 2011.

41. Baylor NW, Midthun K. Regulation and testing of vaccines. In: Plotkin S, Orenstein WA, Offit PA, editors. *Vaccines*, 5th edn. Philadelphia: Saunders, 2008, pp. 1611–27.
42. Ellis RW. Technologies for making new vaccines. In: Plotkin S, Orenstein WA, Offit PA, editors. *Vaccines*, 5th edn. Philadelphia: Saunders, 2008, pp. 1335–55.
43. Jackson LA, Yu O, Nelson J, Belongia EA, Hambidge SJ, Baxter R, *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**: 4912–6.
44. Bernstein DI, Edwards KM, Dekker CL, Belshe R, Talbot HK, Graham IL, *et al.* Effects of adjuvants on the safety and immunogenicity of an avian influenza H5N1 vaccine in adults. *J Infect Dis* 2008; **197**: 667–75.
45. Lewis DJ, Huo Z, Barnett S, Kromann I, Giemza R, Galiza E, *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**: e6999.
46. Pool V, Braun MM, Kelso JM, Mootrey G, Chen RT, Yunginger JW, *et al.* Prevalence of anti-gelatin IgE antibodies in people with anaphylaxis after measles-mumps rubella vaccine in the United States. *Pediatrics* 2002; **110**: e71.
47. Zent O, Hennig R. Post-marketing surveillance of immediate allergic reactions: polygeline-based

versus polygeline-free pediatric TBE vaccine. *Vaccine* 2004; **23**: 579–84.
48. Zanoni G, Puccetti A, Dolcino M, Simone R, Peretti A, Ferro A, *et al.* Dextran-specific IgG response in hypersensitivity reactions to measles-mumps-rubella vaccine. *J Allergy Clin Immunol* 2008; **122**: 1233–5.
49. 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* (Madrid) 2010: **38**: 341–3.
50. Parthasarathy A. Contamination of reconstituted multidose measles vaccine vial and toxic shock syndrome in Tamilnadu. *Indian Pediatr* 2008; **45**: 606.
51. Offit PA, Jew RK. Addressing parents' concerns: do vaccines contain harmful preservatives, adjuvants, additives, or residuals? *Pediatrics* 2003; **112**: 1394–7.
52. Hutin YJ, Hauri AM, Armstrong GL. Use of injections in healthcare settings worldwide, 2000: literature review and regional estimates. *BMJ* 2003; **327**(7423): 1075.
53. Miller D, Wadsworth J, Diamond J, Ross E. Pertussis vaccine and whooping cough as risk factors in acute neurologic illness and deaths in young children. *Dev Biol Stand* 1985; **61**: 389–94.
54. Schonberger LB, Bregman DJ, Sullivan-Bolyai JZ, Keenlyside RA, Ziegler DW, Retailliau HF, *et al.* Guillain-Barre syndrome following vaccination in the National Influenza Immunization Program, United States, 1976–1977. *Am J Epidemiol* 1979; **110**: 105–23.
55. Centers for Disease Control and Prevention. Preliminary Results: Surveillance for Guillain–Barre Syndrome After Receipt of Influenza A (H1N1) 2009 Monovalent Vaccine–United States, 2009–2010. *MMWR Morb Mortal Wkly Rep* 2010; **59**: 657–61.
56. Centers for Disease Control and Prevention. Poliomyelitis prevention in the United States: introduction of a sequential vaccination schedule of inactivated poliovirus vaccine followed by oral poliovirus vaccine. Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep* 1997; **46**: 1–25.
57. McGoldrick MD, Bailie GR. Nonnarcotic analgesics: prevalence and estimated economic impact of toxicities. *Ann Pharmacother* 1997; **31**: 221–7.
58. Marcuse EK, Wentz KR. The NCES reconsidered: summary of a 1989 workshop. National Childhood Encephalopathy Study. *Vaccine* 1990; **8**: 531–5.
59. Alderslade R, Bellman MH, Rawson NSB, Ross EM, Miller DL. *The National Childhood Encephalopathy Study. Whooping Cough: Reports from the Committee on the Safety of Medicines and the Joint Committee on Vaccination and Immunisation*. Department of Health and Social Security. London: HM Stationary Office, 1981, pp. 79–169.
60. Wentz KR, Marcuse EK. Diphtheria-tetanus-pertussis vaccine and serious neurologic illness: an updated review of the epidemiologic evidence. *Pediatrics* 1991; **87**: 287–97.
61. Berkovic SF, Harkin L, McMahon JM, Pelekanos JT, Zuberi SM, Wirrell EC, *et al.* De-novo mutations of the sodium channel gene SCN1A in alleged vaccine encephalopathy: a retrospective study. *Lancet Neurol* 2006; **5**: 488–92.
62. Neustadt RE, Fineberg HV. *The Swine Flu Affair: Decision-Making on a Slippery Disease*. Washington, DC: US Government Printing Office, 1978.
63. Kurland LT, Wiederholt WC, Kirkpatrick JW, Potter HG, Armstrong P. Swine influenza vaccine and Guillain-Barre syndrome. Epidemic or artifact? *Arch Neurol* 1985; **42**: 1089–90.
64. Langmuir AD, Bregman DJ, Kurland LT, Nathanson N, Victor M. An epidemiologic and clinical evaluation of Guillain–Barre syndrome reported in association with the administration of swine influenza vaccines. *Am J Epidemiol* 1984; **119**: 841–79.
65. Safranek TJ, Lawrence DN, Kurland LT, Culver DH, Wiederholt WC, Hayner NS, *et al.* Reassessment of the association between Guillain–Barre syndrome and receipt of swine influenza vaccine in 1976–1977: results of a two-state study. Expert Neurology Group. *Am J Epidemiol* 1991; **133**: 940–51.
66. Murphy TV, Gargiullo PM, Massoudi MS, Nelson DB, Jumaan AO, Okoro CA, *et al.* Intussusception among infants given an oral rotavirus vaccine. *N Engl J Med* 2001; **344**: 564–72.
67. Kramarz P, France EK, DeStefano F, Black SB, Shinefield H, Ward JI, *et al.* Population-based study of rotavirus vaccination and intussusception. *Pediatr Infect Dis J* 2001; **20**: 410–6.
68. 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.
69. Hall AJ. Ecological studies and debate on rotavirus vaccine and intussusception. *Lancet* 2001; **358** (9289): 1197–8.

70. Verstraeten T, DeStefano F, Chen RT, Miller E. Vaccine safety surveillance using large linked databases: opportunities, hazards and proposed guidelines. *Expert Rev Vaccines* 2003; **2**: 21–9.
71. Salmon DA, Sapsin JW, Teret S, Jacobs RF, Thompson JW, Ryan K, *et al.* Public health and the politics of school immunization requirements. *Am J Public Health* 2005; **95**: 778–83.
72. John TJ, Samuel R. Herd immunity and herd effect: new insights and definitions. *Eur J Epidemiol* 2000; **16**: 601–6.
73. Hardin G. The tragedy of the commons. *Science* 1968; **162**: 1243–8.
74. Fine PE. Methodological issues in the evaluation and monitoring of vaccine safety. *Ann N Y Acad Sci* 1995; **754**: 300–8.
75. Bauch CT, Earn DJ. Vaccination and the theory of games. *Proc Natl Acad Sci U S A* 2004; **101**: 13391–4.
76. Manfredi P, Posta PD, d'Onofrio A, Salinelli E, Centrone F, Meo C, *et al.* Optimal vaccination choice, vaccination games, and rational exemption: an appraisal. *Vaccine* 2009; **28**: 98–109.
77. Nathanson N, Langmuir AD. The Cutter incident. *Am J Hygiene* 1963; **78**: 16–8.
78. Offit PA. *The Cutter Incident: How America's First Polio Vaccine Led to the Growing Vaccine Crisis*. New Haven: Yale University Press, 2005.
79. Immunization Safety Review Committee. *SV40 Contamination of Polio Vaccine and Cancer*. Washington, DC: The National Academies Press, 2002.
80. Food and Drug Administration. Bovine-derived materials: Agency letters to manufacturers of FDA-regulated products. *Fed Reg* 1994; **59**: 44591–4.
81. Victoria JG, Wang C, Jones MS, Jaing C, McLoughlin K, Gardner S, *et al.* Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. *J Virol* 2010; **84**: 6033–40.
82. World Health Organization. *Procedure for Assessing the Acceptability, in Principle, of Vaccines for Purchase by United Nations Agencies*. Geneva: World Health Organization, 2006. Report No. WHO/IVB/05.19.
83. Centers for Disease Control and Prevention. Poliomyelitis prevention in the United States: updated recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep* 2000; **49**: 1–25.
84. Anonymous. Two MMR vaccines withdrawn. *Lancet* 1992; **340**: 922.
85. Schmitt HJ, Just M, Neiss A. Withdrawal of a mumps vaccine: reasons and impacts. *Eur J Pediatr* 1993; **152**: 387–8.
86. Kimura M, Kuno-Sakai H, Yamazaki S, Yamada A, Hishiyama M, Kamiya H, *et al.* Adverse events associated with MMR vaccines in Japan. *Acta Paediatr Jpn* 1996; **38**: 205–11.
87. Lathrop SL, Ball R, Haber P, Mootrey GT, Braun MM, Shadomy SV, *et al.* Adverse event reports following vaccination for Lyme disease: December 1998–July 2000. *Vaccine* 2002; **20**: 1603–8.
88. Guillaume L, Bath PA. A content analysis of mass media sources in relation to the MMR vaccine scare. *Health Informatics J* 2008; **14**: 323–34.
89. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine* 2010; **28**: 1709–16.
90. Yawata M. Japan's troubles with measles-mumps-rubella vaccine. *Lancet* 1994; **343**: 105–6.
91. Lloyd JC, Chen RT. The Urabe mumps vaccine: Lessons in adverse event surveillance and response. *Pharmacoepidemiol Drug Saf* 1996; **5**: 1996.
92. Evans G, Levine EM, Saindon EH. Legal issues. In: Plotkin S, Orenstein WA, Offit PA, eds. *Vaccines*, 5th edn. Philadelphia: Saunders, 2008, pp. 1651–75.
93. Stewart AM. When vaccine injury claims go to court. *N Engl J Med* 2009; **360**: 2498–500.
94. Institute of Medicine. *Review of Adverse Effects of Vaccines*. http://www.iom.edu/Activities/Research/VaccineAdvEffectReview.aspx (accessed July 2011).
95. 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**: 597–605.
96. Centers for Disease Control and Prevention. Public Health Service recommendations on smallpox vaccination. *MMWR Morb Mortal Wkly Rep* 1971; **20**: 339–45.
97. Bottiger M. The elimination of polio in the Scandinavian countries. *Public Health Rev* 1993; **21**: 27–33.
98. Bonnet MC, Dutta A. World wide experience with inactivated poliovirus vaccine. *Vaccine* 2008; **26**: 4978–83.
99. 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**: 459–66.
100. Henderson DA. Countering the posteradication threat of smallpox and polio. *Clin Infect Dis* 2002; **34**: 79–83.

101. Howson CP, Fineberg HV. Adverse events following pertussis and rubella vaccines. Summary of a report of the Institute of Medicine. *JAMA* 1992; **267**: 392–6.
102. 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**: 1602–5.
103. Farrington CP, Pugh S, Colville A, Flower A, Nash J, Morgan-Capner P, *et al.* A new method for active surveillance of adverse events from diphtheria/tetanus/pertussis and measles/mumps/rubella vaccines. *Lancet* 1995; **345**: 567–9.
104. Dourado I, Cunha S, Teixeira MG, Farrington CP, Melo A, Lucena R, *et al.* Outbreak of aseptic meningitis associated with mass vaccination with a urabe-containing measles-mumps-rubella vaccine: implications for immunization programs. *Am J Epidemiol* 2000; **151**: 524–30.
105. Ki M, Park T, Yi SG, Oh JK, Choi B. Risk analysis of aseptic meningitis after measles-mumps-rubella vaccination in Korean children by using a case-crossover design. *Am J Epidemiol* 2003; **157**: 158–65.
106. Mutsch M, Zhou W, Rhodes P, Bopp M, Chen RT, Linder T, *et al.* Use of the inactivated intranasal influenza vaccine and the risk of Bell's palsy in Switzerland. *N Engl J Med* 2004; **350**: 896–903.
107. Jacobsen SJ, Ackerson BK, Sy LS, Tran TN, Jones TL, Yao JF, *et al.* Observational safety study of febrile convulsion following first dose MMRV vaccination in a managed care setting. *Vaccine* 2009; **27**: 4656–61.
108. Stetler HC, Orenstein WA, Bart KJ, Brink EW, Brennan JP, Hinman AR. History of convulsions and use of pertussis vaccine. *J Pediatr* 1985; **107**: 175–9.
109. Burns CC, Campagnoli R, Shaw J, Vincent A, Jorba J, Kew O. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *J Virol* 2009; **83**: 9957–69.
110. Monath TP, Lee CK, Julander JG, Brown A, Beasley DW, Watts DM, *et al.* Inactivated yellow fever 17D vaccine: development and nonclinical safety, immunogenicity and protective activity. *Vaccine* 2010; **28**: 3827–40.
111. Barrett AD, Beasley DW. Development pathway for biodefense vaccines. *Vaccine* 2009; **27** (Suppl 4): D2–D7.
112. Pichichero ME. Acellular pertussis vaccines. Towards an improved safety profile. *Drug Saf* 1996; **15**: 311–24.
113. Goldenthal KL, McVittie LD. The clinical testing of preventive vaccines. In: Mathieu M, editor. *Biologic Development: A Regulatory Overview*. Waltham, MA: Paraxel, 1993, pp. 119–30.
114. Francis T, Korns R, Voight R. An evaluation of the 1954 poliomyelitis vaccine trials: summary report. *Am J Public Health* 1955; **45**: 1–63.
115. Heaton PM, Ciarlet M. Vaccines: the pentavalent rotavirus vaccine: discovery to licensure and beyond. *Clin Infect Dis* 2007; **45**: 1618–24.
116. Heyse JF, Kuter BJ, Dallas MJ, Heaton P. Evaluating the safety of a rotavirus vaccine: the REST of the story. *Clin Trials* 2008; **5**: 131–9.
117. Dos Santos BA, Ranieri TS, Bercini M, Schermann MT, Famer S, Mohrdieck R, *et al.* An evaluation of the adverse reaction potential of three measles-mumps-rubella combination vaccines. *Rev Panam Salud Publica* 2002; **12**: 240–6.
118. Chen RT. Safety of acellular pertussis vaccine: follow-up studies. *Dev Biol Stand* 1997; **89**: 373–5.
119. Bonhoeffer J, Kohl K, Chen R, Duclos P, Heijbel H, Heininger U, *et al.* The Brighton Collaboration: addressing the need for standardized case definitions of adverse events following immunization (AEFI). *Vaccine* 2002; **21**: 298–302.
120. Heijbel H, Ciofi degli Atti MC, Harzer E, Liese J, Preziosi MP, Rasmussen F, *et al.* Hypotonic hyporesponsive episodes in eight pertussis vaccine studies. *Dev Biol Stand* 1997; **89**: 101–3.
121. Fine PE, Chen RT. Confounding in studies of adverse reactions to vaccines. *Am J Epidemiol* 1992; **136**: 121–35.
122. 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**: 220–32.
123. 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**: 457–61.
124. Hall AJ, Cutts FT. Lessons from measles vaccination in developing countries. *BMJ* 1993; **307**: 1294–5.
125. Cutts FT, Fine PE. Caution–mortality ratios ahead. *Lancet* 2003; **361**: 2169–70.
126. Black S. Global Vaccine Safety DataNet meeting. *Expert Rev Vaccines* 2008; **7**: 15–20.

127. Black S, Mulholland K, Halsey NA. Developing sustainable funding for vaccine safety infrastructure. *Vaccine* 2010; **28**: 1133–4.
128. Zanoni G, Dung NTM, Tridente G. *Fourth Report on the Green Channel Activities and the Monitoring System of Reactions to Vaccines in the Veneto Region: Analysis of the Three Year Period 1997–1999, General Summary 1992–99*. University of Verona, Italy Immunology Section, 2000.
129. 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**: 549–52.
130. Gold MS. Hypotonic-hyporesponsive episodes following pertussis vaccination: a cause for concern? *Drug Saf* 2002; **25**: 85–90.
130a. LaRussa PS, Edwards KM, Dekker CL, Klein NP, Halsey NA, Marchant C, *et al*. Understanding the role of human variation in vaccine adverse events: the Clinical Immunization Safety Assessment Network. *Pediatrics* 2011 May; **127** Suppl 1: S65–73.
131. Van Dam CN, Syed S, Eron JJ, Ostrander M, Engler RJ, Damon I, *et al.* Severe postvaccinia encephalitis with acute disseminated encephalomyelitis: recovery with early intravenous immunoglobulin, high-dose steroids, and vaccinia immunoglobulin. *Clin Infect Dis* 2009; **48**: e47–e49.
132. Skowronski DM, Strauss B, De Serres G, MacDonald D, Marion SA, Naus M, *et al.* Oculo-respiratory syndrome: a new influenza vaccine-associated adverse event? *Clin Infect Dis* 2003; **36**: 705–13.
133. Quinn P, Gold M, Royle J, Buttery J, Richmond P, McIntyre P, *et al.* Recurrence of extensive swelling reactions following DTPa or dTpa vaccine in children 4–6 years old. *Vaccine* 2010; **29**: 4230–7.
134. Baraff LJ, Cody CL, Cherry JD. DTP-associated reactions: an analysis by injection site, manufacturer, prior reactions, and dose. *Pediatrics* 1984; **73**: 31–6.
135. Baraff LJ, Manclark CR, Cherry JD, Christenson P, Marcy SM. Analyses of adverse reactions to diphtheria and tetanus toxoids and pertussis vaccine by vaccine lot, endotoxin content, pertussis vaccine potency and percentage of mouse weight gain. *Pediatr Infect Dis J* 1989; **8**: 502–7.
136. Centers for Disease Control and Prevention. Notice to Readers: Manufacturer's Recall of Human Rabies Vaccine–April 2, 2004. *MMWR Recomm Rep* 2004; **53** (Dispatch): 1–2.
137. Chen RT, Pool V, Takahashi H, Weniger BG, Patel B. Combination vaccines: postlicensure safety evaluation. *Clin Infect Dis* 2001; **33** (Suppl 4): S327–S333.
138. Wise RP, Kiminyo KP, Salive ME. Hair loss after routine immunizations. *JAMA* 1997; **278**: 1176–8.
139. Centers for Disease Control and Prevention. Intussusception among recipients of rotavirus vaccine–United States, 1998–1999. *MMWR Morb Mortal Wkly Rep* 1999; **48**: 577–81.
140. Chen RT. Surveillance for Vaccine Adverse Events and Public Health Disease: Similarities and Differences. *Pharmacoepidemiol Drug Saf* 1996; **5** (Suppl 1): S45.
141. Division of Immunization. Vaccine-associated adverse events in Canada, 1992 Report. *Can Commun Dis Rep* 1995; **21**: 117–28.
142. Chen RT, Rastogi SC, Mullen JR, Hayes SW, Cochi SL, Donlon JA, *et al.* The Vaccine Adverse Event Reporting System (VAERS). *Vaccine* 1994; **12**: 542–50.
143. Kohl K, Bonhoeffer J, Chen R. The Brighton Collaboration: enhancing comparability of vaccine safety data. *Pharmacoepidemiol Drug Saf* 2003; **12**: 335–40.
144. Miller E, Goldacre M, Pugh S, Colville A, Farrington P, Flower A, *et al.* Risk of aseptic meningitis after measles, mumps, and rubella vaccine in UK children. *Lancet* 1993; **341**: 979–82.
145. Schwartz B, Weniger BG, Chen RT. *Vaccine identification standards initiative*. Paper presented at: The 32nd National Immunization Conference, 1998 July 21–24, *Atlanta*, Georgia, USA.
146. Halsey NA. An update on pediatric immunization. *Adv Pediatr Infect Dis* 1995; **10**: 187–225.
147. Thompson WW, Price C, Goodson B, Shay DK, Benson P, Hinrichsen VL, *et al.* Early thimerosal exposure and neuropsychological outcomes at 7 to 10 years. *N Engl J Med* 2007; **357**: 1281–92.
148. Price CS, Thompson WW, Goodson B, Weintraub ES, Croen LA, Hinrichsen VL, *et al.* Prenatal and infant exposure to thimerosal from vaccines and immunoglobulins and risk of autism. *Pediatrics* 2010; **126**: 656–64.
149. Velentgas P, Bohn RL, Brown JS, Chan KA, Gladowski P, Holick CN, *et al.* A distributed research network model for post-marketing safety studies: the Meningococcal Vaccine Study. *Pharmacoepidemiol Drug Saf* 2008; **17**: 1226–34.
150. Lasky T, Terracciano GJ, Magder L, Koski CL, Ballesteros M, Nash D, *et al.* The Guillain–Barre syndrome and the 1992–1993 and 1993–1994 influenza vaccines. *N Engl J Med* 1998; **339**: 1797–802.

151. Miller D, Wadsworth J, Diamond J, Ross E. Pertussis vaccine and whooping cough as risk factors in acute neurological illness and death in young children. *Dev Biol Stand* 1985; **61**: 389–94.
152. Laboratory Centre for Disease Control. Report of the working group on the possible relationship between hepatitis B vaccination and the chronic fatigue syndrome. *Can Med Assoc J* 1993; **149**: 314–19.
152a. Bardage C, Persson I, Ortqvist A, Bergman U, Ludvigsson JF, Granath F. Neurological and autoimmune disorders after vaccination against pandemic influenza A (H1N1) with a monovalent adjuvanted vaccine: population based cohort study in Stockholm, Sweden. *BMJ* 2011; **12**: 343:d5956. doi:10.1136/bmj.d5956. [Epub ahead of print].
153. Bernier RH, Frank JA, Jr., Dondero TJ, Jr., Turner P. Diphtheria-tetanus toxoids-pertussis vaccination and sudden infant deaths in Tennessee. *J Pediatr* 1982; **101**: 419–21.
154. Thompson NP, Montgomery SM, Pounder RE, Wakefield AJ. Is measles vaccination a risk factor for inflammatory bowel disease? *Lancet* 1995; **345**: 1071–4.
155. Expanded Programme on Immunization. Lack of evidence that hepatitis B vaccine causes multiple sclerosis. *Wkly Epidemiol Rec* 1997; **72**: 149–56.
156. Barlow WE, Davis RL, Glasser JW, Rhodes PH, Thompson RS, Mullooly JP, *et al.* The risk of seizures after receipt of whole-cell pertussis or measles, mumps, and rubella vaccine. *N Engl J Med* 2001; **345**: 656–61.
157. Donahue JG, Kieke BA, Yih WK, Berger NR, McCauley JS, Baggs J, *et al.* Varicella vaccination and ischemic stroke in children: is there an association? *Pediatrics* 2009; **123**: e228–e234.
158. DeStefano F, Gu D, Kramarz P, Truman BI, Iademarco MF, Mullooly JP, *et al.* Childhood vaccinations and risk of asthma. *Pediatr Infect Dis J* 2002; **21**: 498–504.
159. Kramarz P, DeStefano F, Gargiullo PM, Davis RL, Chen RT, Mullooly JP, *et al.* Influenza vaccination in children with asthma in health maintenance organizations. Vaccine Safety Datalink Team. *Vaccine* 2000; **18**: 2288–94.
160. DeStefano F, Mullooly JP, Okoro CA, Chen RT, Marcy SM, Ward JI, *et al.* Childhood vaccinations, vaccination timing, and risk of type 1 diabetes mellitus. *Pediatrics* 2001; **108**: E112.
161. Ascherio A, Zhang SM, Hernan MA, Olek MJ, Coplan PM, Brodovicz K, *et al.* Hepatitis B vaccination and the risk of multiple sclerosis. *N Engl J Med* 2001; **344**: 327–32.
162. Black C, Kaye JA, Jick H. MMR vaccine and idiopathic thrombocytopaenic purpura. *Br J Clin Pharmacol* 2003; **55**: 107–11.
163. Mortimer EA, Jones P, Adelson L. DTP and SIDS—Letter. *Pediatr Infect Dis* 1983; **2**: 492–493.
164. Andrews NJ. Statistical assessment of the association between vaccination and rare adverse events post-licensure. *Vaccine* 2001; **20** (Suppl 1): S49–53; discussion S45–8, S49–S53.
165. Kramarz P, DeStefano F, Gargiullo PM, Chen RT, Lieu TA, Davis RL, *et al.* Does influenza vaccination prevent asthma exacerbations in children? *J Pediatr* 2001; **138**: 306–10.
166. Miller E, Waight P, Farrington CP, Andrews N, Stowe J, Taylor B. Idiopathic thrombocytopenic purpura and MMR vaccine. *Arch Dis Child* 2001; **84**: 227–9.
167. France EK, Glanz JM, Xu S, Davis RL, Black SB, Shinefield HR, *et al.* Safety of the trivalent inactivated influenza vaccine among children: a population-based study. *Arch Pediatr Adolesc Med* 2004; **158**: 1031–6.
168. Hambidge SJ, Glanz JM, France EK, McClure D, Xu S, Yamasaki K, *et al.* Safety of trivalent inactivated influenza vaccine in children 6 to 23 months old. *JAMA* 2006; **296**: 1990–7.
169. Ward KN, Bryant NJ, Andrews NJ, Bowley JS, Ohrling A, Verity CM, *et al.* Risk of serious neurologic disease after immunization of young children in Britain and Ireland. *Pediatrics* 2007; **120**: 314–21.
170. France EK, Glanz J, Xu S, Hambidge S, Yamasaki K, Black SB, *et al.* Risk of immune thrombocytopenic purpura after measles-mumps-rubella immunization in children. *Pediatrics* 2008; **121**: e687–e692.
171. Stowe J, Andrews N, Taylor B, Miller E. No evidence of an increase of bacterial and viral infections following measles, mumps and rubella vaccine. *Vaccine* 2009; **27**: 1422–5.
172. Naleway AL, Belongia EA, Donahue JG, Kieke BA, Glanz JM. Risk of immune hemolytic anemia in children following immunization. *Vaccine* 2009; **27**: 7394–7.
173. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 1995; **51**: 228–35.
174. Glanz JM, McClure DL, Xu S, Hambidge SJ, Lee M, Kolczak MS, *et al.* Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations. *J Clin Epidemiol* 2006; **59**: 808–18.

175. Rennels MB. The rotavirus vaccine story: a clinical investigator's view. *Pediatrics* 2000; **106**: 123–5.
176. 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**: 364–76.
177. Ray WA, Griffin MR. Re: "Confounding in studies of adverse reactions to vaccines". *Am J Epidemiol* 1994; **139**: 229–30.
178. Ellenberg SS. Safety considerations for new vaccine development. *Pharmacoepidemiol Drug Saf* 2001; **10**: 411–5.
179. Lesko SM, Mitchell AA. The safety of acetaminophen and ibuprofen among children younger than two years old. *Pediatrics* 1999; **104**: e39.
180. Scott TF, Bonanno DE. Reactions to live-measles-virus vaccine in children previously inoculated with killed-virus vaccine. *N Engl J Med* 1967; **277**: 248–50.
181. Kristensen I, Aaby P, Jensen H. Routine vaccinations and child survival: follow up study in Guinea-Bissau, West Africa. *BMJ* 2000; **321**: 1435–8.
182. Bennett JV, Cutts FT, Katz SL. Edmonston-Zagreb measles vaccine: A good vaccine with an image problem. *Pediatrics* 1999; **104**: 1123–4.
183. CIOMS Working Group VII. *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.
184. Rock EP, Molloy VJ, Humphrey JS. GCP data quality for early clinical development. *Clin Cancer Res* 2010; **16**: 1756–63.
185. Mgone CS, Salami W. EDCTP: a genuine north-south partnership. *Trop Med Int Health* 2009; **14**: 1327–8.
186. *International Workshop: Harmonization of Reporting of Adverse Events Following Vaccination*. Rockville, MD: Food and Drug Administration/CBER, 1993.
187. Edwards IR, Lindquist M, Hugman B. Global Drug Surveillance: The WHO Programme for International Drug Monitoring. In: Strom BL, ed. *Pharmacoepidemiology*, 4th edn. Philadelphia: John Wiley & Sons, Ltd, 2007, pp. 161–83.
188. Lankinen KS, Pastila S, Kilpi T, Nohynek H, Makela PH, Olin P. Vaccinovigilance in Europe—need for timeliness, standardization and resources. *Bull World Health Organ* 2004; **82**: 828–35.
189. Zanoni G, Berra P, Lucchi I, Ferro A, O'Flanagan D, Levy-Bruhl D, *et al*. Vaccine adverse event monitoring systems across the European Union countries: time for unifying efforts. *Vaccine* 2009; **27**: 3376–84.
190. Takahashi H. Need for improved vaccine safety surveillance. *Vaccine* 2000; **19**: 1004.
191. Mansoor O, Pillans PI. Vaccine adverse events reported in New Zealand 1990–5. *N Z Med J* 1997; **110**: 270–2.
192. 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**: 1185–94.
193. Mahajan D, Menzies R, Roomiani I, Lawrence G. Supplementary report: surveillance of adverse events following immunisation among children aged less than seven years in Australia, 1 January to 30 June 2009. *Commun Dis Intell* 2010; **34**: 49–53.
194. Monteiro SA, Takano OA, Waldman EA. Surveillance for adverse events after DTwP/Hib vaccination in Brazil: sensitivity and factors associated with reporting. *Vaccine* 2010; **28**: 3127–33.
195. Public Health Agency of Canada. *Canadian Adverse Events Following Immunization Surveillance System (CAEFISS)*, Feb 7 2008. http://www.phac-aspc.gc.ca/im/vs-sv/caefiss-eng.php (accessed June 2010).
196. Galindo Santana BM, Galindo Sardina MA, Perez RA. [Adverse reaction surveillance system for vaccination in the Republic of Cuba]. *Rev Cubana Med Trop* 1999; **51**: 194–200.
197. Andersen MM, Ronne T. Side-effects with Japanese encephalitis vaccine. *Lancet* 1991; **337**: 1044.
198. Sokhey J. Adverse events following immunization: 1990. *Indian Pediatr* 1991; **28**: 593–607.
199. Squarcione S, Vellucci L. Adverse reactions following immunization in Italy in the years 1991–93. *Ig Mod* 1996; **105**: 1419–31.
200. Fescharek R, Arras-Reiter C, Arens ER, Quast U, Maass G. [Oral vaccines against poliomyelitis and vaccination-related paralytic poliomyelitis in Germany. Do we need a new immunization strategy?] *Wien Med Wochenschr* 1997; **147**: 456–61.
201. Labadie J, van Grootheest AC. [Adverse events following vaccination reported to the Netherlands Pharmacovigilance Center Lareb in 2004–2006]. *Ned Tijdschr Geneeskd* 2007; **151**: 2738–42.
202. Tatley MV, Kunac DL, McNicholas A, Zhou L, Ballantyne S, Ashton J, *et al*. The Intensive Vaccines Monitoring Programme (IVMP): an electronic system to monitor vaccine safety in New Zealand. *Vaccine* 2008; **26**: 2746–52.
203. Schumacher Z, Bourquin C, Heininger U. Surveillance for adverse events following immuni-

zation (AEFI) in Switzerland 1991–2001. *Vaccine* 2010; **28**: 4059–64.

204. Galea SA, Sweet A, Beninger P, Steinberg SP, LaRussa PS, Gershon AA, *et al.* The safety profile of varicella vaccine: a 10-year review. *J Infect Dis* 2008; **197** (*Suppl 2*): S165–S169.
205. Hussey GD, Lasser ML, Reekie WD. The costs and benefits of a vaccination programme for *Haemophilus influenzae* type B disease. *S Afr Med J* 1995; **85**: 20–5.
206. 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**: 170–7.
207. National childhood vaccine injury act. Requirements for permanent vaccination records and for reporting of selected events after vaccination. National Vaccine Program, Centers for Disease Control. *Ohio Med* 1988; **84**: 754–6.
208. Haber P, Rashidee A, Zhou W, English R. Web-Based Reporting: 10 Months Experience in the VAERS. *Pharmacoepidemiol Drug Saf* 2003; **12**: S1–S189.
209. Centers for Disease Control and Prevention. Smallpox Vaccine Adverse Events Monitoring and Response System for the first stage of the smallpox vaccination program. *MMWR Morb Mortal Wkly Rep* 2003; **52**: 88–9, 99.
210. 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*, Aug 21, 2009. http://www.cdc.gov/h1n1flu/vaccination/safety_planning.htm (accessed June 2010).
211. Hinrichsen VL, Kruskal B, O'Brien MA, Lieu TA, Platt R. Using electronic medical records to enhance detection and reporting of vaccine adverse events. *J Am Med Inform Assoc* 2007; **14**: 731–5.
212. Rashidee AH, Chen RT, Lugg M, Black S. Integrating vaccine safety initiatives in immunization registry/systems. *Pharmacoepidemiol Drug Saf* 2002; **11**: S195.
213. 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**: E97.
214. Centers for Disease Control and Prevention. Adverse Events Associated with 17D-Derived Yellow Fever Vaccination United States 2001–2002. *MMWR Morb Mortal Wkly Rep* 2002; **51**: 989–93.
215. Centers for Disease Control and Prevention. Cardiac adverse events following smallpox vaccination—United States, 2003. *MMWR Morb Mortal Wkly Rep* 2003; **52**: 248–50.
216. Centers for Disease Control and Prevention. Surveillance for adverse events associated with anthrax vaccination—U.S. Department of Defense, 1998–2000. *MMWR Morb Mortal Wkly Rep* 2000; **49**: 341–5.
217. Zhou W, Pool V, Iskander JK, English-Bullard R, Ball R, Wise RP, *et al.* Surveillance for safety after immunization: Vaccine Adverse Event Reporting System (VAERS)—United States, 1991–2001. *MMWR Surveill Summ* 2003; **52**: 1–24.

217a. Vellozzi C, Broder KR, Haber P, Guh A, Nguyen M, Cano M, *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**: 7248–55.

218. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf* 1999; **20**: 109–17.
219. Scheifele D, Duval B, De Serres G, Skowronski DM. Unique roles of a data and safety monitoring board in vaccine safety trials with compressed timelines and urgent implications. *Control Clin Trials* 2003; **24**: 99–104.
220. Sauve LJ, Bettinger J, Scheifele D, Halperin S, Vaudry W, Law B. Postvaccination thrombocytopenia in Canada. *Pediatr Infect Dis J* 2010; **29**: 559–61.
221. 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**: 178–85.
222. Vermeer-de Bondt PE, Wesselo C, Dzaferagic A, Phaff TAJ. *Adverse Events Following Immunisation under the National Vaccination Programme of the Netherlands*. Part VIII-reports in 2001. Bilthoven: RIVM report 000001007, 2003.

222a. Williams SE, Pahud BA, Vellozzi C, Donofrio PD, Dekker CL, Halsey N, *et al.* Causality assessment of serious neurologic adverse events following 2009 H1N1 vaccination. *Vaccine* 2011 Sep 2. [Epub ahead of print].

223. Salisbury D, Ramsay M, Noakes K. Immunisation Against Infectious Disease, 3rd edn. London: The Stationery Office under license from the Department of Health, 2006.
224. World Health Organization. *WHO Collaborating Centre for International Drug Monitoring. The Uppsala*

Monitoring Centre, 2011. http://www.who-umc.org/ (accessed July 2011).

225. World Health Organization. *Surveillance of Adverse Events Following Immunization*. Geneva: WHO/EPI/TRAM/93.02 Rev. 1, 1997.
226. World Health Organization. Vaccine supply and quality: surveillance of adverse events following immunization. *Weekly Epidemiology Rec* 1996; **71**: 237–42.
227. Domingo C, Niedrig M. Safety of 17D derived yellow fever vaccines. *Expert Opin Drug Saf* 2009; **8**: 211–21.
228. Hesseling AC, Marais BJ, Gie RP, Schaaf HS, Fine PE, Godfrey-Faussett P, *et al.* The risk of disseminated Bacille Calmette-Guerin (BCG) disease in HIV-infected children. *Vaccine* 2007; **25**: 14–8.
229. Fenichel GM, Lane DA, Livengood JR, Horwitz SJ, Menkes JH, Schwartz JF. Adverse events following immunization: assessing probability of causation. *Pediatr Neurol* 1989; **5**: 287–90.
230. Wassilak SGF, Sokhey J. *Monitoring of Adverse Events Following Immunization in the Expanded Programme on Immunization*. Geneva: World Health Organization/EPI/GEN, 1991, Report No.: 91.2.
231. Goodman M, Lamm SH, Bellman MH. Temporal relationship modeling: DTP or DT immunizations and infantile spasms. *Vaccine* 1998; **16**: 225–31.
232. Health Council of the Netherlands. *Committee on Adverse Reactions to Vaccinations. Adverse Reactions to Vaccines in the National Immunization Programme in 2002–2003*. The Hague: Health Council of the Netherlands, 2006, Report no. 2006/14.
233. Brito GS. *System of Investigation and Notification of Adverse Events Following Immunization, Preliminary Report*. State of Sao Paolo, Brazil: Health Department, 1991.
234. Laboratory Centre for Disease Control. *Proceedings of a Workshop on the Standardization of Definitions for Post-marketing Surveillance of Adverse Vaccine Reactions, Ottawa, Canada, October 30–31*, 1991. Ottawa, Canada: Laboratory Centre for Disease Control, 1992, pp. 1–17.
235. Braun MM, Terracciano G, Salive ME, Blumberg DA, Vermeer-de Bondt PE, Heijbel H, *et al.* Report of a US public health service workshop on hypotonic-hyporesponsive episode (HHE) after pertussis immunization. *Pediatrics* 1998; **102**: E52.
236. 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.
237. 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.
238. Wood RA, Berger M, Dreskin SC, Setse R, Engler RJ, Dekker CL, *et al.* An algorithm for treatment of patients with hypersensitivity reactions after vaccines. *Pediatrics* 2008; **122**: e771–e777.
239. 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.
240. Pulendran B, Miller J, Querec TD, Akondy R, Moseley N, Laur O, *et al.* Case of yellow fever vaccine-associated viscerotropic disease with prolonged viremia, robust adaptive immune responses, and polymorphisms in CCR5 and RANTES genes. *J Infect Dis* 2008; **198**: 500–7.
241. Rennels MB. Extensive swelling reactions occurring after booster doses of diphtheria-tetanus-acellular pertussis vaccines. *Semin Pediatr Infect Dis* 2003; **14**: 196–8.
242. Sim I, Sanders GD, McDonald KM. Evidence-based practice for mere mortals: the role of informatics and health services research. *J Gen Intern Med* 2002; **17**: 302–8.
243. Kramer MS, Lane DA. Causal propositions in clinical research and practice. *J Clin Epidemiol* 1992; **45**: 639–49.
244. Hill AB. The environment and disease: association or causation. *Proc Roy Soc Med* 1965; **58**: 295–300.
245. Kramer MS. Difficulties in assessing the adverse effects of drugs. *Br J Clin Pharmacol* 1981; **11**: 105–110.
246. Jones JK. Determining causation from case reports. In: Strom BL, ed. *Pharmacoepidemiology*, 3rd edn. Sussex, UK: John Wiley & Sons, 2000, pp. 525–38.
247. Edsall G, Elliott MW, Peebles TC, Eldred MC. Excessive use of tetanus toxoid boosters. *JAMA* 1967; **202**: 111–3.
248. Henderson DA, Witte JJ, Morris L, *et al.* Paralytic disease associated with oral polio vaccines. *JAMA* 1964; **190**: 153–60.
249. Centers for Disease Control and Prevention. Rubella vaccination during pregnancy—United States, 1971–1988. *MMWR Morb Mortal Wkly Rep* 1989; **38**: 289–93.
250. Chen RT, Glasser JW, Rhodes PH, Davis RL, Barlow WE, Thompson RS, *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**: 765–73.
251. Cherry JD. Pertussis and the vaccine controversy. In: Roote RK, Griffiss JM, Warren KS, *et al.,* eds. *Immunization*. New York: Churchill Livingstone, 1989, pp. 47–63.
252. World Health Organization. *Adverse Events Following Immunization (AEFI): Causality Assessment*. WHO Department of Vaccines and Biologicals, 2010.
253. Lane DA. A probabilist's view of causality assessment. *Drug Inform J* 1984; **18**: 323–30.
254. Aagaard L, Hansen EH. Information about ADRs explored by pharmacovigilance approaches: a qualitative review of studies on antibiotics, SSRIs and NSAIDs. *BMC Clin Pharmacol* 2009; **9**: 4.
255. Retailliau HF, Curtis AC, Storr G, Caesar G, Eddins DL, Hattwick MA. Illness after influenza vaccination reported through a nationwide surveillance system, 1976–1977. *Am J Epidemiol* 1980; **111**: 270–8.
256. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, *et al.* A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol* 1998; **54**: 315–21.
257. Kulldorff M. Prospective time periodic geographic disease surveillance using a scan statistic. *J Roy Stat Soc Series A* 2001; **164**: 61–72.
258. Naus J, Wallenstein S. Temporal surveillance using scan statistics. *Stat Med* 2006; **25**: 311–24.
259. 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**: 924–5.
260. Martin M, Tsai TF, Cropp B, Chang GJ, Holmes DA, Tseng J, *et al.* Fever and multisystem organ failure associated with 17D-204 yellow fever vaccination: a report of four cases. *Lancet* 2001; **358**: 98–104.
261. Martin M, Weld LH, Tsai TF, Mootrey GT, Chen RT, Niu M, *et al.* Advanced age a risk factor for illness temporally associated with yellow Fever vaccination. *Emerg Infect Dis* 2001; **7**: 945–51.
262. Monath TP. Suspected yellow fever vaccine-associated viscerotropic adverse events (1973 and 1978), United States. *Am J Trop Med Hyg* 2010; **82**: 919–21.
263. Centers for Disease Control and Prevention. Update: Guillain–Barre syndrome among recipients of Menactra meningococcal conjugate vaccine—United States, October 2005–February 2006. *MMWR Morb Mortal Wkly Rep* 2006; **55**: 364–6.
264. Neale T. ACIP: Warning for Guillain–Barré syndrome lifted. *MedPage Today* Jun 24, 2010.
265. Commonwealth Chief Medical Officer. *Seasonal Flu Vaccine Remains Suspended for young children without risk factors—Advice from the Chief Medical Officer*, June 1, 2010. http://www.health.gov.au/internet/main/publishing.nsf/Content/mr-yr10-dept-dept010610.htm (accessed July 2011).
265a. Immunization Safety Office Update: Vaccines and Febrile Seizures. http://www.cdc.gov/vaccines/recs/acip/slides-jun11.htm#general (accessed September 2011).
266. Carson JL, Strom BL, Maislin G. Screening for unknown effects of newly marketed drugs. In: Strom BL, ed. *Pharmacoepidemiology*, 2nd edn. Sussex, UK: John Wiley & Sons, 1994, pp. 431–47.
267. Walker AM. Pattern recognition in health insurance claims databases. *Pharmacoepidemiol Drug Saf* 2001; **10**: 393–7.
268. Ball R. Methods for ensuring vaccine safety. *Expert Rev Vaccines* 2002; **1**: 161–8.
269. Zhou W, Pool V, DeStefano F, Iskander JK, Haber P, Chen RT. A potential signal of Bell's palsy after parenteral inactivated influenza vaccines: reports to the Vaccine Adverse Event Reporting System (VAERS) –United States, 1991–2001. *Pharmacoepidemiol Drug Saf* 2004; **13**: 505–10.
270. Haber P, Iskander J, English-Bullard R. Use of proportional reporting rate ratio in monitoring vaccine adverse event reports. *Pharmacoepidemiol Drug Saf* 2002; **11**: S229.
271. 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**: 4627–34.
272. 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.
273. 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**: 551–62.
274. 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**: 457–60.
275. 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**: 483–6.

276. 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**: 601–9.
277. Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K, *et al.* Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 2007; **45** (10 Suppl 2): S89–S95.
278. Greene SK, Kulldorff M, Lewis EM, Li R, Yin R, Weintraub ES, *et al.* Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project. *Am J Epidemiol* 2010; **171**: 177–88.
279. Davis RL, Kolczak M, Lewis E, Nordin J, Goodman M, Shay DK, *et al.* Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology* 2005; **16**: 336–41.
280. 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.
281. Klein NP, Fireman B, Yih WK, Lewis E, Kulldorff M, Ray P, *et al.* Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. *Pediatrics* 2010: **126**: e1–8.
282. Walker AM. Signal detection for vaccine side effects that have not been specified in advance. *Pharmacoepidemiol Drug Saf* 2010; **19**: 311–7.
283. Balinska MA. L'affaire hepatite B en France. *Espirit* 2001; **276**: 34–48.
284. Whittembury A, Ramirez G, Hernandez H, Ropero AM, Waterman S, Ticona M, *et al.* Viscerotropic disease following yellow fever vaccination in Peru. *Vaccine* 2009; **27**: 5974–81.
285. 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**: 1106–12.
286. Uhari M, Rantala H, Niemela M. Cluster of childhood Guillain-Barre cases after an oral poliovaccine campaign. *Lancet* 1989; **2** (8660): 440–1.
287. Berg SW, Mitchell BS, Hanson RK, Olafson RP, Williams RP, Tueller JE, *et al.* Systemic reactions in U.S. Marine Corps personnel who received Japanese encephalitis vaccine. *Clin Infect Dis* 1997; **24**: 265–6.
288. 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**: 2287–92.
289. Yergeau A, Alain L, Pless R, Robert Y. Adverse events temporally associated with meningococcal vaccines. *Can Med Assoc J* 1996; **1154**: 503–7.
290. De WP, Deceuninck G, Ouakki M, Boulianne N, De SG, Danzig L. Analysis of mortality following a mass immunization campaign with serogroup C meningococcal conjugate vaccine: methodological difficulties and imperfect solutions. *Vaccine* 2009; **27**: 3223–7.
291. da Silveira CM, Salisbury DM, de Quadros CA. Measles vaccination and Guillain–Barre syndrome. *Lancet* 1997; **349**: 14–6.
292. Rantala H, Cherry JD, Shields WD, Uhari M. Epidemiology of Guillain–Barre syndrome in children: relationship of oral polio vaccine administration to occurrence. *J Pediatr* 1994; **124**: 220–3.
293. Sexton K, McNicholas A, Galloway Y, Radke S, Kieft C, Stehr-Green P, *et al.* Henoch–Schonlein purpura and meningococcal B vaccination. *Arch Dis Child* 2009; **94**: 224–6.
294. Clements CJ. Mass psychogenic illness after vaccination. *Drug Saf* 2003; **26**: 599–604.
295. Huang WT, Hsu CC, Lee PI, Chuang JH. Mass psychogenic illness in nationwide in-school vaccination for pandemic influenza A(H1N1) 2009, Taiwan, November 2009–January 2010. *Euro Surveill* 2010; **15**: 19575.
296. McNicholas A, Galloway Y, Stehr-Green P, Reid S, Radke S, Sexton K, *et al.* Post-marketing safety monitoring of a new group B meningococcal vaccine in New Zealand, 2004–2006. *Hum Vaccin* 2007; **3**: 196–204.
297. Johansen K, Nicoll A, Ciancio BC, Kramarz P. Pandemic influenza A(H1N1) 2009 vaccines in the European Union. *Euro Surveill* 2009; **14**: 19361.
298. Huang WT, Chuang JH, Kuo SH. Monitoring the safety of pandemic H1N1 vaccine. *Lancet* 2010; **375**: 1164.
299. Black S, Eskola J, Siegrist CA, Halsey N, MacDonald N, Law B, *et al.* Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic H1N1 influenza vaccines. *Lancet* 2009; **374**: 2115–22.
300. Pool V, Iskander J. Safety of influenza vaccination during pregnancy. *Am J Obstet Gynecol* 2006; **194**: 1200.
301. 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**: 529–35.

302. 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**: 2908–17.
303. 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.
304. 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**: 88–90.
305. Braun MM, Patriarca PA, Ellenberg SS. Syncope after immunization. *Arch Pediatr Adolesc Med* 1997; **151**: 255–9.
306. 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**: e353–9.
307. Wise RP, Salive ME, Braun MM, Mootrey GT, Seward JF, Rider LG, *et al.* Postlicensure safety surveillance for varicella vaccine. *JAMA* 2000; **284**: 1271–9.
308. Rastogi SC, Wise RP. Exploring associations between vaccine lot potencies and human safety experience. *Pharmacoepidemiol Drug Saf* 1996; **5**: S75.
309. Ellenberg SS, Chen RT. The complicated task of monitoring vaccine safety. *Public Health Rep* 1997; **112**: 10–20.
310. Dayan G, Iskander JK, English-Bullard R, Glasser J, Chen RT. Tracking the lifecycle of vaccines and vaccine lots via the Vaccine Adverse Event Reporting System (VAERS). *Pharmacoepidemiol Drug Saf* 2002; **11**: S117.
311. 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**: 333–6.
312. Barwick R. History of thymoma and yellow fever vaccination. *Lancet* 2004; **364**: 936.
313. Chaves SS, Haber P, Walton K, Wise RP, Izurieta HS, Schmid DS, *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.
314. Centers for Disease Control and Prevention (CDC). Syncope after vaccination—United States, January 2005–July 2007. *MMWR Morb Mortal Wkly Rep* 2008; **57**: 457–60.
315. Rosenthal S, Chen R. The reporting sensitivities of two passive surveillance systems for vaccine adverse events. *Am J Public Health* 1995; **85**: 1706–9.
316. 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**: 1006–12.
317. Chen RT, Kent JH, Rhodes PH, *et al.* Investigation of a possible association between influenza vaccination and Guillain–Barre syndrome in the United States, 1990–1991. *Postmarketing Surveillance* 1992; **6**: 5–6.
318. Finney DJ. The detection of adverse reactions to therapeutic drugs. *Stat Med* 1982; **1**: 153–61.
319. Bonnet MC, Dutta A, Weinberger C, Plotkin SA. Mumps vaccine virus strains and aseptic meningitis. *Vaccine* 2006; **24**: 7037–45.
320. De Serres G, Grenier JL, Toth E, Menard S, Roussel R, Tremblay M, *et al.* The clinical spectrum of the oculo-respiratory syndrome after influenza vaccination. *Vaccine* 2003; **21**: 2354–61.
321. De Serres G., Toth E, Menard S, Grenier JL, Roussel R, Tremblay M, *et al.* Oculo-respiratory syndrome after influenza vaccination: trends over four influenza seasons. *Vaccine* 2005; **23**: 3726–32.
322. 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**: 1898–904.
323. Niu MT, Rhodes P, Salive M, Lively T, Davis DM, Black S, *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**: 503–10.
324. Slade BA, Leidel L, Vellozzi C, Woo EJ, Hua W, Sutherland A, *et al.* Postlicensure safety surveillance for quadrivalent human papillomavirus recombinant vaccine. *JAMA* 2009; **302**: 750–7.
325. Vellozzi C, Broder KR, Haber P, Guh A, Nguyen M, Cano M, *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**: 7248–55.
326. 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.
327. Begg NT, Gill ON, White JM. COVER (cover of vaccination evaluated rapidly): description of the England and Wales scheme. *Public Health* 1989; **103**: 81–9.
328. Freeman VA, DeFriese GH. The challenge and potential of childhood immunization registries. *Annu Rev Public Health* 2003; **24**: 227–46.
329. Decker MD, Edwards KM. Acellular pertussis vaccines. *Pediatr Clin North Am* 2000; **47**: 309–35.
330. Sano J, Chaitman BR, Swindle J, Frey SE. Electrocardiography screening for cardiotoxicity after modified Vaccinia Ankara vaccination. *Am J Med* 2009; **122**: 79–84.
331. Patriarca PA, Laender F, Palmeira G, Oliveira MJ, Lima FJ, Dantes MC, *et al.* Randomised trial of alternative formulations of oral poliovaccine in Brazil. *Lancet* 1988; **1** (8583): 429–33.
332. Bhargava I, Chhaparwal BC, Phadke MA, Irani SF, Chakladhar BK, Maheshwari CP. Reactogenecity of indigenously produced measles vaccine. *Indian Pediatr* 1996; **33**: 827–31.
333. 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.
334. Booy R, Taylor SA, Dobson SR, Isaacs D, Sleight G, Aitken S, *et al.* Immunogenicity and safety of PRP-T conjugate vaccine given according to the British accelerated immunisation schedule. *Arch Dis Child* 1992; **67**: 475–8.
335. Deforest A, Long SS, Lischner HW, Girone JA, Clark JL, Srinivasan R, *et al.* Simultaneous administration of measles-mumps-rubella vaccine with booster doses of diphtheria-tetanus-pertussis and poliovirus vaccines. *Pediatrics* 1988; **81**: 237–46.
336. 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**: 1399–406.
337. Fine PEM, Sterne J. High titer measles vaccine before nine months of age: Implications for child survival. *Consultation on Studies Involving High Titer Measles Vaccines Before Nine Months of Age*, Atlanta, GA, June 16–17, 1992. Working paper no. 13.
338. Aaby P, Samb B, Simondon F, Knudsen K, Seck AM, Bennett J, *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**: 326–30.
339. Anon. Expanded programme on immunization (EPI). Safety of high titer measles vaccines. *Wkly Epidemiol Rec* 1992; **67**: 357–61.
340. Black SB, Shinefield HR. b-CAPSA I Haemophilus influenzae, type b, capsular polysaccharide vaccine safety. *Pediatrics* 1987; **79**: 321–5.
341. Meekison W, Hutcheon M, Guasparini R. Postmarketing surveillance of adverse events following PROHIBIT vaccine in British Columbia. *Can Dis Wkly Rep* 1989; **15**: 143–5.
342. 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**: 555–61.
343. Centers for Disease Control and Prevention. Unintentional administration of varicella virus vaccine—United States, 1996. *MMWR Morb Mortal Wkly Rep* 1996; **45**: 1017–8.
344. Black S, Shinefield H, Ray P, Lewis E, Hansen J, Schwalbe J, *et al.* Postmarketing evaluation of the safety and effectiveness of varicella vaccine. *Pediatr Infect Dis J* 1999; **18**: 1041–6.
345. Wong C, Krashin J, Rue-Cover A, Saraiya M, Unger E, Calugar A, *et al.* Invasive and in situ cervical cancer reported to the vaccine adverse event reporting system (VAERS). *J Womens Health (Larchmt)* 2010; **19**: 365–70.
346. Haber P, Patel M, Izurieta HS, Baggs J, Gargiullo P, Weintraub E, *et al.* Postlicensure monitoring of intussusception after RotaTeq vaccination in the United States, February 1, 2006, to September 25, 2007. *Pediatrics* 2008; **121**: 1206–12.
347. A JE, Simonsen L, Viboud C, Steiner C, Patel MM, Curns AT, *et al.* Trends in intussusception hospitalizations among US infants, 1993–2004: implications for monitoring the safety of the new rotavirus vaccination program. *Pediatrics* 2008; **121**: e1125–e1132.
348. Anon. Rotavirus vaccine and intussusception: report from an expert consultation. *Wkly Epidemiol Rec* 2011; **86**: 317–21.
349. Centers for Disease Control and Prevention. Inactivated Japanese encephalitis virus vaccine. Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep* 1993; **42**: 1–15.
350. Liu ZL, Hennessy S, Strom BL, Tsai TF, Wan CM, Tang SC, *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**: 1366–9.
351. Heijbel H, Chen RT, Dahlquist G. Cumulative incidence of childhood-onset IDDM is unaffected by pertussis immunization. *Diabetes Care* 1997; **20**: 173–5.
352. Solberg LK. *DTP Vaccination, Visit to Child Health Center and Sudden Infant Death Syndrome (SIDS): Evaluation of DTP Vaccination. Report to the Oslo Health Council 1985*. Besthesda, MD: NIH Library Translation, 1985.
353. Bouvier-Colle MH, Flahaut A, Messiah A, Jougla E, Hatton F. Sudden infant death and immunization: an extensive epidemiological approach to the problem in France—winter 1986. *Int J Epidemiol* 1989; **18**: 121–6.
354. Mitchell EA, Stewart AW, Clements M. Immunisation and the sudden infant death syndrome. New Zealand Cot Death Study Group. *Arch Dis Child* 1995; **73**: 498–501.
355. Gale JL, Thapa PB, Wassilak SG, Bobo JK, Mendelman PM, Foy HM. Risk of serious acute neurological illness after immunization with diphtheria-tetanus-pertussis vaccine. A population-based case-control study. *JAMA* 1994; **271**: 37–41.
356. Fukumi H, Kimura M, Kato S, Muranaka S, Hayashi H, Kawana R, *et al.* [Clinical report on safety and antigenicity of 23-valent pneumococcal polysaccharide vaccine]. *Kansenshogaku Zasshi* 1984; **58**: 495–511.
357. 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**: 368–71.
358. Farrington CP, Miller E, Taylor B. MMR and autism: further evidence against a causal association. *Vaccine* 2001; **19**: 3632–5.
359. Fombonne E, Chakrabarti S. No evidence for a new variant of measles-mumps-rubella-induced autism. *Pediatrics* 2001; **108**: E58.
360. 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**: 945–51.
361. 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**: 618–23.
362. DeStefano F. The Vaccine Safety Datalink project. *Pharmacoepidemiol Drug Saf* 2001; **10**: 403–6.
363. Mayes R. The origins, development, and passage of Medicare's revolutionary prospective payment system. *J Hist Med Allied Sci* 2007; **62**: 21–55.
364. Baggs J, Gee J, Lewis E, Fowler G, Benson P, Lieu T, *et al.* The Vaccine Safety Datalink: A model for monitoring immunization safety. *Pediatrics* 2011; **127** (Suppl 1): S45–53.
365. Ray P, Black S, Shinefield H, Dillon A, Schwalbe J, Holmes S, *et al.* Risk of chronic arthropathy among women after rubella vaccination. Vaccine Safety Datalink Team. *JAMA* 1997; **278**: 551–6.
366. Yu O, Bohlke K, Hanson CA, Delaney K, Rees TG, Zavitkovsky A, *et al.* Hepatitis B vaccine and risk of autoimmune thyroid disease: a Vaccine Safety Datalink study. *Pharmacoepidemiol Drug Saf* 2007; **16**: 736–45.
367. Klein NP, Massolo ML, Greene J, Dekker CL, Black S, Escobar GJ. Risk factors for developing apnea after immunization in the neonatal intensive care unit. *Pediatrics* 2008; **121**: 463–9.
368. Davis RL, Marcuse E, Black S, Shinefield H, Givens B, Schwalbe J, *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**: 767–71.
369. Lin ND, Kleinman K, Chan KA, Soumerai S, Mehta J, Mullooly JP, *et al.* Multiple vaccinations and the risk of medically attended fever. *Vaccine* 2010; **28**: 4169–74.
370. Belongia EA, Irving SA, Shui IM, Kulldorff M, Lewis E, Yin R, *et al.* Real-time surveillance to assess risk of intussusception and other adverse events after pentavalent, bovine-derived rotavirus vaccine. *Pediatr Infect Dis J* 2010; **29**: 1–5.
371. Orenstein WA, Rodewald LE, Hinman AR, Schuchat A. Immunizations in the United States. In: Plotkin SA, Orenstein WA, Offit PA, eds. *Vaccines*, 5th edn. Philadelphia: Saunders, 2008, pp. 1479–509.
372. Verstraeten T, Jumaan AO, Mullooly JP, Seward JF, Izurieta HS, DeStefano F, *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**: e98–103.
373. Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. *Stat Med* 2010; **29**: 284–95.

374. 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**: 1641–5.
375. Rothman KJ. *Causal Inference*. Chestnut Hill, MA: Epidemiology Resources Inc., 1988.
376. Andrews N, Miller E, Waight P, Farrington P, Crowcroft N, Stowe J, *et al.* Does oral polio vaccine cause intussusception in infants? Evidence from a sequence of three self-controlled cases series studies in the United Kingdom. *Eur J Epidemiol* 2001; **17**: 701–6.
377. Kaye JA, Mar Melero-Montes M, Jick H. Mumps, measles, and rubella vaccine and the incidence of autism recorded by general practitioners: a time trend analysis. *BMJ* 2001; **322**: 460–3.
378. Roberts JD, Roos LL, Poffenroth LA, Hassard TH, Bebchuk JD, Carter AO, *et al.* Surveillance of vaccine-related adverse events in the first year of life: a Manitoba cohort study. *J Clin Epidemiol* 1996; **49**: 51–8.
379. Ali M, Canh DG, Clemens JD, Park JK, von Seidlein L, Thiem VD, *et al.* The vaccine data link in Nha Trang, Vietnam: a progress report on the implementation of a database to detect adverse events related to vaccinations. *Vaccine* 2003; **21**: 1681–6.
380. Griffin MR, Ray WA, Fought RL, Foster MA, Hays A, Schaffner W. Monitoring the safety of childhood immunizations: methods of linking and augmenting computerized data bases for epidemiologic studies. *Am J Prev Med* 1988; **4** (2 Suppl): 5–13.
381. Mullooly J, Drew L, DeStefano F, Chen R, Okoro K, Swint E, *et al.* Quality of HMO vaccination databases used to monitor childhood vaccine safety. Vaccine Safety DataLink Team. *Am J Epidemiol* 1999; **149**: 186–94.
382. Wilton R, Pennisi A. Evaluating the accuracy of transcribed computer-stored immunization data. *Pediatrics* 1994; **94**: 902–6.
383. Boulianne N, Hemon YA, Mawhinney T, Strong D, Gemmill I, Dobson S, *et al.* National eligible, due, and overdue guidelines for immunization registries: draft recommendations from the Canadian Immunization Registry Network, Data Standards Task Group. *Can Commun Dis Rep* 2004; **30**: 53–9.
384. Hull BP, Lawrence GL, MacIntyre CR, McIntyre PB. Immunisation coverage in Australia corrected for under-reporting to the Australian Childhood Immunisation Register. *Aust N Z J Public Health* 2003; **27**: 533–8.
385. 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**: 254–6.
386. Street RL, Jr. Information-giving in medical consultations: the influence of patients' communicative styles and personal characteristics. *Soc Sci Med* 1991; **32**: 541–8.
387. Anon. Global Advisory Committee on Vaccine Safety, 3–4 December 2009. *Wkly Epidemiol Rec* 2010; **85**: 29–33.
388. Bohlke K, Davis RL, Marcy SM, Braun MM, DeStefano F, Black SB, *et al.* Risk of anaphylaxis after vaccination of children and adolescents. *Pediatrics* 2003; **112**: 815–20.
389. DeStefano F, Verstraeten T, Jackson LA, Okoro CA, Benson P, Black SB, *et al.* Vaccinations and risk of central nervous system demyelinating diseases in adults. *Arch Neurol* 2003; **60**: 504–9.
390. Poser CM. Neurological complications of swine influenza vaccination. *Acta Neurol Scand* 1982; **66**: 413–31.
391. Bardage C, Olsson J, Persson I, Bergman U, Ortqvist A. Vaccine vigilance: safety of pandemic flu vaccine in the population of Stockholm, Sweden. *Pharmacoepidemiol Drug Saf* 2010; **19**: S328.
392. 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**: 1363–7.
393. Greene SK, Kulldorff M, Lewis EM, Li R, Yin R, Weintraub ES, *et al.* Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project. *Am J Epidemiol* 2010; **171**: 177–88.
394. Rothman KJ, Greenland S. *Modern Epidemiology*. New York: Lippincott Williams & Wilkins, 1998.
395. Szklo M, Nieto M. Basic study designs in analytical epidemiology. In: Szklo M, Nieto M, eds. *Epidemiology: Beyond the Basics*, 2nd edn. Sudbury, MA: Jones and Bartlett Publishers, 2007, pp. 3–43.
396. Black C, Kaye JA, Jick H. MMR vaccine and idiopathic thrombocytopaenic purpura. *Br J Clin Pharmacol* 2003; **55**: 107–11.
397. Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol* 1996; **143**: 1165–73.
398. Farrington CP, Whitaker HJ. Semiparametric analysis of case series data. *Applied Statistics* 2006; **55**: 553–94.

399. Farrington CP, Whitaker HJ, Hocine MN. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics* 2009; **10**: 3–16.
400. Kramarz P, DeStefano F, Gargiullo PM, Davis RL, Chen RT, Mullooly JP, *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**: 617–23.
401. Mullooly JP, Pearson J, Drew L, Schuler R, Maher J, Gargiullo P, *et al.* Wheezing lower respiratory disease and vaccination of full-term infants. *Pharmacoepidemiol Drug Saf* 2002; **11**: 21–30.
402. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006; **25**: 1768–97.
403. 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**: 650–6.
404. Xu S, Zhang L, Nelson J, Zeng C, Mullooly J, McClure D, *et al.* Identifying optimal risk windows for self-controlled case series studies of vaccine safety. *Stat Med* 2010; **30**: 142–52.
405. Omer SB, Pan WK, Halsey NA, Stokley S, Moulton LH, Navar AM, *et al.* Nonmedical exemptions to school immunization requirements: secular trends and association of state policies with pertussis incidence. *JAMA* 2006; **296**: 1757–63.
406. Omer SB, Enger KS, Moulton LH, Halsey NA, Stokley S, Salmon DA. Geographic clustering of nonmedical exemptions to school immunization requirements and associations with geographic clustering of pertussis. *Am J Epidemiol* 2008; **168**: 1389–96.
407. Glanz JM, McClure DL, Magid DJ, Daley MF, France EK, Salmon DA, *et al.* Parental refusal of pertussis vaccination is associated with an increased risk of pertussis infection in children. *Pediatrics* 2009; **123**: 1446–51.
408. 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**: 375–84.
409. Platt R, Madre L, Reynolds R, Tilson H. Active drug safety surveillance: a tool to improve public health. *Pharmacoepidemiol Drug Saf* 2008; **17**: 1175–82.

CHAPTER 27

Epidemiologic Studies of Medical Devices: Methodologic Considerations for Implantable Devices

Danica Marinac-Dabic,[1] *Sharon-Lise Normand,*[2] *Art Sedrakyan,*[3] *and Thomas Gross*[1]

[1] Office of Surveillance and Biometrics, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, USA

[2] Department of Biostatistics, Harvard School of Public Health and Harvard Medical School, Boston, MA, USA

[3] Weill Cornell Medical College, New York, NY, USA

Introduction

Recent decades have seen a dramatic growth in medical device technology worldwide.[1] It has been estimated that the US medical device market exceeded $100 billion in 2008, representing roughly 42% of the worldwide device market.[2] In addition to the new generations of conventional devices, groundbreaking innovations in the areas of nanotechnology, telemedicine, minimally invasive/non-invasive procedures, and sophisticated health information technology continue to offer new diagnostic and therapeutic options to patients and clinicians.

Naturally, the expansion of such a diverse technology has created a demand for methodologic approaches for optimal study and timely information on device benefits and risks. Hence, the impact of device innovation in the context of modern medicine has sharpened the focus on medical device epidemiology, a relatively new discipline within the field of epidemiology. Medical device epidemiology is not only well suited to study the extent of utilization of medical devices, but also to study utilization patterns and identification of risk for certain outcomes in defined populations. Furthermore, an attractive feature of modern device epidemiology involves implementation of techniques to integrate information available from the growing body of heterogeneous data. Operating at the very intersection of scientific knowledge and health care, it is the practice of epidemiology that ensure consistently reliable approaches to combine and update information in order to maximize quality, minimize bias, and to reduce the uncertainty in understanding risk–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 27.1). 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: (i) recognized in the official National Formulary, or United States Pharmacopeia or any supplement of them; (ii) 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 (iii) intended to affect the

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

Table 27.1 International definitions of medical devices

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.
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.

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, Canada, Australia, and Japan are slightly different (Table 27.1).[3–9]

The Global Harmonization Task Force, 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 medical device.[9]

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. Depending upon the complexity of the devices and their intended use, there are variety of issues to be evaluated which determine the type and the depth of the premarket data necessary for approval. Based on the varying level of benefit–risk evidence needed over different categories, devices are classified into three regulatory classes. Class I devices present minimal potential for harm to the patient and require neither clinical testing nor special controls (e.g., standards) to establish reasonable assurance of safety and effectiveness (e.g., elastic bandages, gloves, manual surgical instruments). Class II devices by definition are higher-risk (e.g., infusion pumps, diagnostic ultrasound machines) and as such are subject to additional regulatory control including special controls and, for certain device groups, clinical testing to demonstrate safety and effectiveness. 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 The effectiveness and safety of these devices have to be determined based on a 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."[10–12]

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 European Union, manufacturers of devices of Classes II and III, as well as devices of Class I with either measuring function or sterility requirements, must submit to the regulator (competent authority): (i) a Declaration of Conformity to the appropriate 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. In Australia, 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. In Japan, all devices above Class II must obtain a central government license for market entry.[4]

The common regulatory theme underpinning the 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 Class II and Class III) because of their significant public health impact, high risk for adverse events, and uncertainties surrounding the long-term exposure.

Implantable medical devices comprise an important device category in the very heterogeneous world of medical devices. As true of other devices, implantables share characteristics that distinguish them not only from other devices, but also from regulated drugs. Table 27.2 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

Table 27.2 Characteristics of medical devices as compared to drugs (US Regulations)

Characteristic	Characteristics of medical devices as compared to drugs	
	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	NDC codes
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
	Non-compliance	Same
	Interactions	More prominent
	Malfunctions (manufacturing)	Drug quality problems
	Environmental hazards	Same
	Toxic/allergic	More prominent
	Packaging defects	Same
	Software glitches	
	Poor maintenance	

device category, in general, has longer product life cycles although incremental changes do occur. Implantable devices may be comprised of multiple components (such as a total hip implant) or single components (such as a pacemaker lead). Exposures to such devices are typically chronic (exception being temporary implantables, such as inferior vena cava filters), with the onset of exposure clearly defined at 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 in the case of ruptured breast implants, or if absorbed over time such as with biologically-based dermal fillers or barrier adhesion devices).

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, allergic, autoimmune effects). Additional hazards may be related to human factors (e.g., proper programming of pacemakers) and interactions (e.g., 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 problem to be addressed by pharmacoepidemiologic research

Diffusion to clinical practice and utilization

Introduction and adoption of new medical devices can often induce a breakthrough transformation of clinical practice.[13] Diffusion of medical device technology to practice is influenced by a large number of factors including device complexity, a relative advantage when compared with similar available treatments, positions of opinion-leading organizations, health-care reimbursement decisions, commercial competition, evidence-based guidelines, publication of key research papers, regulatory actions, medical liability issues, and legislative environment.[14] Accordingly, 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 usual healthcare settings when compared to patient experience in premarket clinical trails. A thoughtful epidemiologic assessment of factors that can potentially influence the adoption of new technology can inform regulatory science (both premarket and postmarket), help the development of clinical guidelines and policies, and can significantly shape national reimbursements strategies.[15]

Risk–benefit profile in a real-word setting

When randomized clinical trials (RCTs) for new devices are designed, often elite investigators are selected to participate. For example, in the RCTs of devices used in surgery the operators are typically early adopters, highly skilled, and quick learners, which impacts 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.[16] The rate of learning (the shape of the learning curve) and interaction with other variables for an average surgeon is very difficult to gauge from the RCT that includes only elite 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, is it possible to estimate the learning curve effect 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.[17–19] Other studies have shown a volume threshold for procedures, above which increasing volume is no longer associated with improved outcomes.[20,21] Lastly, others 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) (Resnic FS, 2009, personal communication). 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 learning curve include 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.[22] This has been the case with stenting of carotid arteries by operators from varying specialties (e.g., radiologists, cardiologists, and neurosurgeons).

By the nature of their design, RCTs involve select, non-representative populations (see Chapter 3). A number of studies have highlighted the disparities in disease prevalence, progression, and health outcomes of medical device technology in subgroups of the population.[23,24] Premarket medical device trials often lack sufficient representation of important patient populations (women, children, 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.[25,26] The public health utility of observational studies has been increasing [25] with advances in medical device data capture in medical records, electronic databases, and prospective registries, and the development of innovative analytical tools using observational data.[26]

Recent increase in national interest in comparative effectiveness facilitated by the American Recovery and Reinvestment Act of 2009 and subsequent health-care reform legislation,[27] has begun focusing the national attention on building methods and infrastructure for emerging Comparative Effectiveness Research (CER, see Chapter 32). CER, as defined by the US Institute of Medicine (IOM), is a "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 markers to make informed decisions that will improve health care at both the individual and population levels." CER is vital for the improvement of the health-care quality, better regulatory decisions, and thoughtful guidelines for clinicians and patients.[28]

In recent years, several other countries have established agencies to evaluate health technologies and inform health-care policy decisions.[29] These organizations are different in terms of structures, methods, and processes but in all of them CER is an effort that is aiming to address the needs of payers, patients, clinical professionals, and policymakers. A 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.[30,31] 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, 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 the preclinical testing and premarket clinical trials. FDA's postmarket attention is therefore increasingly directed towards ensuring that studies of sufficient size and length of follow-up are conducted in the postmarket setting to better predict and illustrate the problems occurring long term.[32] 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, United Kingdom, Australia, Canada, and other countries).[33–36]

Methodologic problems to be solved by pharmacoepidemiologic research

Evidence generation for implantable medical devices requires taking into account unique issues that typically do not arise when evaluating benefits and risks of drugs. The key issues are the interaction of device, operator, and the interventional area in which the device is being used. Furthermore, the device design, its complexity, and its specific mechanical characteristics, can be as important as the clinical details such as type of the lesion being treated, severity of the disease, and concomitant

therapy provided. In the commonly used pharmacoepidemiologic research databases these details are often only partially available and sometimes are missing.

Challenges in individual patient exposure assessment

In general, there are two major obstacles, among others, to accurately determine individual patient exposure to a medical device. First, unlike pharmaceuticals where the National Drug Code (NDC) Directory has been established and broadly used, there is no common medical device nomenclature utilized by all stakeholders. The routine capture of device information, such as the name of manufacturer, brand, or model, linked to device group terms, would allow proper identification and easier data management. Second, 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 brand. Thus, capturing complete device exposure information is far more complex for devices than it is for drugs. Significant efforts toward consolidating all existing medical device terminologies, developing unique device identification codes, and promoting their routine documentation in medical records will continue. Once adopted, such a robust, widely incorporated medical device nomenclature will significantly enhance safety surveillance and epidemiologic studies of medical devices.[37]

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 will have to include incorporation of unique device identifiers (UDIs) into data systems, including electronic health records, and routine documentation of device use and patient problems associated with that use. Currently, in the less than ideal national surveillance environment, the 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, market data, etc. In addition, these sources differ in their level of device-specific granularity (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.[37]

Challenges in comparative studies

Epidemiologic/ population-based research relies on non-experimental rather than RCT data to develop evidence about safety and effectiveness of medical products. While there is a recognition of the limitations related to observational data as compared to RCTs,[38] we also need to recognize two facts.

First, methodologic concepts of randomization, allocation concealment, masking/ blinding, withdrawal/follow-up, and intention-to-treat analyses that are recognized as critical components of high methodologic validity, are applicable to pharmaceutical studies but only some can be reasonably applied to evidence development for medical devices.[31]

Second, observational data will frequently complement randomized data rather than replace randomized data. Aside from recognized limitations of clinical trials (e.g., select study subjects, small sample sizes, short duration), the need for data observed in routine clinical practice arises because of learning-curve issues, product modifications, and risks of adverse events. These features are very real issues that affect the safety and effectiveness of medical devices but are only partially applicable to pharmaceuticals.

Addressing issues of sample size, and real-world performance

Typically, premarket clinical trials are powered first and foremost for effectiveness outcomes. Powering the RCTs for less common or rare but serious side effects is not feasible in most instances (see Chapter 4). RCT of devices, due to small sample size and participant selection, often lack *generalizability,* which is defined as the extension of research findings and conclusions from a study conducted on a sample population to the population at large.[39]

Systematic reviews with meta-analyses (see Chapter 40) are observational studies that attempt to capitalize on the detailed data collection within each study. We support the use of systematic reviews as one mechanism to address the small study problems of the RCTs. Systematic reviews with meta-analysis are based on the premise that most of the individual RCTs 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 observational 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 the devices in real-world populations and are based on a solid scientific knowledge accumulated in recent decades. With large observational studies, one can evaluate relevant subgroup effects as well as rare safety and effectiveness endpoints that can not be captures by RCTs.

Ensuring comparability of study groups

Cohort designs offer the opportunity to create comparable groups of patients exposed to devices of interest. Optimally, these designs are based on prospective and consecutive patient enrollment, prospective data collection, and a study that is hypothesis driven. These observational studies should take advantage of statistical adjustments for known and measured confounders and methods that help characterize the impact of residual confounding on results must be incorporated into the analytical strategy (see also Chapter 47).

We have good tools to address unequal distribution in observed patient characteristics (predictors) that is not severely confounded by indication (see Chapter 47). Most of the adjustment techniques deal with imbalances in prognostic 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.[40] Of course, this is an assumption that cannot be tested.

Several analytical methods are available to deal with selection factors and confounding.[40] These methods involve stratification, regression models, or a combination of the two using propensity scores.[41,42] Each approach relies on a set of statistical 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 that of using *instrumental variable-based methods*[43,44] (see Chapter 47).

Currently available solutions

Passive surveillance

Once a device is launched onto the market, 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 United States manufacturers are required to submit reports of device-related deaths, serious injuries, and malfunctions to the FDA. Health-care providers and consumers submit reports voluntarily (through MedWatch).[45] 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 June 2010, MAUDE contained more than 3 million reports. The vast majority of reports in MAUDE (about 200 000 individual per annum) are from manufacturers, with a small percentage from user facilities, voluntary sources, and importers.[46]

Regulatory agencies participating in the Global Harmonization Task Force monitor the safety of devices by reviewing adverse event reports from users, sponsors, other available data sources, and scientific literature.[47]

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; or 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.[48] 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[49,50] (see Chapter 46). 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 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.[51]

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; and (iv) events are generally under-reported and this, in combination with lack of denominator (exposure) data, precludes 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. (See also Chapters 10 and 33.)

Enhanced surveillance

To enhance understanding of clinical issues for medical devices, the Medical Product Safety Network (MedSun) was established to provide national medical device surveillance based on a representative subset of user facilities in the US.[52,53] MedSun currently includes approximately 350 hospitals nationwide. Specialty networks in areas such as electrophysiology devices (HeartNet) and pediatric ICU devices (KidNet) have emerged within MedSun to focus on device-specific issues. Through its ongoing bidirectional interactions, including educational fora, problem solving and posting of reports, and targeted surveys, this enhanced surveillance network helps amplify potential safety signals in real time.

To gain further insight into hospital surveillance dynamics FDA collaborated with a sophisticated tertiary care facility to examine different modes of surveillance (including online incident reporting, computer flags, and methods using discharge claim diagnosis codes). The rate of detection of adverse events varied markedly by mode, from 1.6/1000 discharges for the online mode to 64.6/1000 for the claims mode.[54] Others have examined surveillance in more targeted settings, such as medical and surgical ICUs.[55]

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;[56] (ii) associated with use of recombinant bone morphogenetic protein in cervical spine fusion;[57] and (iii) from MRI-induced interactions in patients with implanted neurological stimulators.[58] Reports received have also spurred the development of significant observational studies elucidating risk factors for meningitis associated with cochlear implants[59] and hemorrhagic complications associated with use of hemostasis devices, including one high-risk device.[60,61]

Active surveillance

When means are available, FDA occasionally uses active surveillance to monitor high-risk devices, such as in a national registry of implanted ventricular-assist devices (VADs) or in *de novo* studies fulfilling postapproval study requirements (see also Part IIIc). The INTERMACS registry attempts to collect complete, detailed, and high quality information on VAD-associated procedures and outcomes, including expected as well as unanticipated device-related adverse events.[62] Designated subsets of these adverse events are adjudicated by external experts.

Another evolving FDA active surveillance effort effectively utilizes the Consumer Product

Safety Commission's (CPSC) National Electronic Injury Surveillance System (NEISS), a sentinel system designed to capture information on consumer-product and other product-related injuries that are seen in hospital emergency departments. FDA collaborated with CPSC to use NEISS to establish the first national estimates of device-related adverse events resulting in visits to emergency departments.[63] More recent efforts focused on examining reasons for device-related visits to emergency departments involving pediatric populations.[64]

To provide more robust surveillance, the FDA launched the Sentinel Initiative in 2007[65] (see Chapter 30). The principal aim of the FDA's Sentinel Initiative is to complement limited FDA postmarket monitoring systems and capabilities with a national, integrated, electronic health-care infrastructure for medical product active surveillance. The surveillance utilizes national, distributed data sources (with populations totaling in the tens of millions), transformed to a common data model(s), against which FDA queries (including active surveillance protocols) can be run, and aggregate data received. In preparation for Sentinel efforts, and to inform them, exploratory device work is being conducted using the Massachusetts statewide coronary intervention registry.[66] In-hospital safety signals (such as myocardial infarction) for recently introduced interventional cardiovascular devices (such as drug-eluting coronary stents) are being explored using an automated computerized safety surveillance system, DELTA (Data Extraction and Longitudinal Trend Analysis system), which was designed to support flexible prospective safety surveillance applicable to a broad range of medical devices.[67,68]

In other countries, national registries of procedures involving implanted medical devices have significantly augmented national surveillance efforts.[33–36] Experience from these registries used for surveillance and observational research will not only provide valuable insights into development of sentinel efforts in the United States, but will also serve as a solid platform for building novel international surveillance infrastructure for regulatory evidence synthesis.

Mandated postmarket studies

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 (PMA) authority under Section 513(a)(3)(C) of the Act; 21 U.S.C. § 360c(a)(3)(C)); 21 C.F.R. Part 814.82(a)(2). A major regulatory/ public health challenge the FDA is facing is to find an appropriate balance for obtaining clinical data premarket to prevent delays in device approval and ensure that only safe and effective devices enter the marketplace. The appropriate postmarket questions that can be answered in a mandated postapproval study include long-term safety and effectiveness, a real-world experience of the device as it enters broader user populations (clinicians and patients), effectiveness of training programs and learning curve effect, and the device performance in certain subgroups of patients not well studied in the premarket clinical trials. Depending on the nature of postmarket questions, a variety of study designs and approaches can be employed. Designing scientifically sound but practical studies and achieving adequate patient and physicians' recruitment rates through adequate minimization of loss to follow-up can be particularly challenging for implantable device studies.

In 2005, the FDA Center for Devices and Radiological Health (CDRH) established the Medical Device Post-Approval Studies (PAS) Program to consolidate the review, tracking, and oversight functions for all medical device postapproval studies imposed by the PMA order. Since then, the CDRH has significantly raised expectations for the quality of new PAS, established a PAS electronic tracking system and a publicly available website posting the study status, begun inspection of selected studies, and instituted routine updates to the Advisory Panels. In parallel, CDRH epidemiologists launched significant efforts to build a robust infrastructure and new innovative epidemiologic methods suitable for PAS.

In addition, Section 522 of Safe Medical Device Act of 1990 and added regulation (21CFR 822) and Section 307 of the FDA Amendments Act of 2007 (FDAAA) (Pub. L. 110-85) allows FDA to require

postmarket surveillance study for Class II and Class III devices that are: (i) intended to be implanted in human body for longer than a year; (ii) life sustaining or life supporting (and used outside of the user facility); (iii) reasonably likely to have serious health consequences if device failure occurred; or (iv) anticipated to have significant use in the pediatric population. Possible study designs vary from detailed review of complaint history and the literature, non-clinical testing, use of registries, observational study designs, and randomized clinical trials. By statute, the duration of these studies is limited to 3 years patient follow-up, with an exception of longer duration in the studies involving pediatric population and there are consequences for not meeting the study requirements.[69]

Registries

A recognition that RCTs cannot fill all the gaps in clinical evidence for implantable devices is not new but has recently regained interest as registries have emerged as powerful tools to harness the full potential of observational studies (see also Chapter 21).

The Agency for Healthcare Research and Quality defines a patient registry for evaluating outcomes as "an organized system that uses 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 a predetermined scientific, clinical, or policy purpose(s)."[70]

Patient registries constitute great infrastructure for conducting large-scale medical device studies. One important value that registries add is their ability to collect data on large numbers of patients, clinical settings, and outcomes over time (see Chapter 21). Depending on the questions to be addressed, the registries can be designed to capture data on the conditions or exposures to a specific device; a type of health-care service delivered (e.g., surgical treatment or diagnostic procedure); or outcome (e.g., adverse event, disorder, or disease).

In the absence of a unique device identification code, the added value of registries for medical devices surveillance include capturing brand/model-specific information crucial for signal identification and comparative effectiveness and safety studies. The complexity and scientific rigor of a registry can vary from those designed to evaluate quality of health care delivered, those specifically established to study 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 the registry data (cohort, case–cohort, case–control, cross sectional, quasiexperimental), both mandated and discretionary.

Major limitations of the registries in general are their voluntary nature and short duration of follow up of patients. For implantable medical devices in particular, the modes of follow-up are critical. Therefore linking the registry data to other data sources is often necessary to ensure the longitudinal data.[71]

Using novel methods these registries can be linked to other databases, including administrative billing data. The linkage of clinically rich procedural and intrahospital data captured by the registry to follow-up data from administrative databases (such as Medicare and Medicaid databases, see below) can substantially augment the value of registries for postmarket assessment of nationally representative samples of implantable devices.[72]

Over time, the regulatory agencies have increased efforts to work with the clinical community, other public health agencies, and academia to foster the development of clinical registries as a valuable postmarket tool for capturing utilization of devices, identifying early signals, and studying postmarket performance of medical technology. Examples of long-standing collaborative efforts include those between FDA/CDRH and professional society databases such as the American College of Cardiology National Cardiovascular Data Registry (NCDR), resulting in one of the largest observational studies on hemostasis devices using NCDR registry data.[60,61] In addition, the FDA has collaborated with Duke University and the Society of Thoracic Surgeons (STS) to study the outcomes of transmyocardial revascularization procedures using the STS Adult Cardio Thoracic Database.[71]

Administrative claims data

The use of administrative databases for epidemiologic research has the strengths of studying large numbers of patients with diverse characteristics and wide varieties of clinical practices, as well as inclusion of longitudinal data from the continuum of clinical care, and good representation of vulnerable populations, leading to increased external validity (generalizability)[73,74] (see Part IIIb). The large number of diverse patients present opportunities to study treatment effect heterogeneity and to advance methods such as high-dimensionality propensity scoring and instrumental variables. With regard to medical devices, limitations of administrative databases include the lack of unique device identifiers, potential inaccuracy of coding of diagnosis, 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. CDRH has used these data to estimate patient characteristics and in-hospital mortality rates associated with aortic valve replacements and pacemaker implantation.[75,76] Others used the data to perform studies on artificial hips.[77–81]

Methods for implantable device outcome evaluation

A methodologic framework for implantable device epidemiology and surveillance involves understanding factors impacting 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 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 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, Zhan *et al.*[81] reported that advanced age, co-morbidities such as heart failure and diabetes, and non-elective 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.[80] The primary reason for the device implant also drives the clinical endpoints.

Surgeon 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 disruptions that are associated with less pain, reduced blood loss, and shorter hospital stays. However, complications can increase if the surgeon is still early in his or her learning curve.[82] Surgical volume of the surgeon and of the center relate to procedural success. Other features of the surgery can impact 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 system may have a porous surface that permits bone to grow into its surface.

Understanding the treatment assignment mechanism

A key principle we adopt to utilizing observational data for surveillance is to view observational studies as approximations to randomized studies. As such, a 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.

With this in mind, assume data have been collected in an observational setting. Let Y_{ij}^{m} denote the mth outcome for the jth subject treated by the ith surgeon. Some of the outcomes may be continuous and some may be discrete. Let T_{ij} denote the device implanted in the jth subject and, for ease of discussion, assume two treatments are of interest so that T_{ij} is a binary-valued variable. The alternative treatment strategy may involve another device or a therapeutic strategy. When comparing one device to one non-device, the definition of the treatment (D) and surgeon (Z) characteristics requires identification of commonly-define factors. In comparing multiple devices with multiple non-devices, treatment and surgeon characteristics within each type of treatment strategy can vary.

In the observational setting, the treatment assignment mechanism is unknown and requires estimation. 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:

$$\text{Logit}(P(T_{ij} = 1)) = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_i + \beta_3 D + \varepsilon_i$$

where ε_i is a surgeon-specific random effect assumed to be symmetric, for example $N(0, \tau^2)$. With more than two devices, extensions using multinomial or nested logit models can be estimated.[83] In the latter case, device characteristics can be included to differentiate among devices within a similar class of devices. 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 scalar summary to validate assumptions required to estimate causal effects. These assumptions relate to unmeasured confounding, positivity of device or treatment assignments, and additive device effects. The positivity assumption[84] 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 for the subject is not defined. The additive treatment effect assumes that the outcome for a subject implanted with device 1 differs by a constant amount from the outcome had the subject been implanted with device 2. The definition of a favorable device effect in this setting requires modification.

Semiparametric estimators, such as matching estimators, weighted estimators, or double-robust estimators that augment weighted estimators with regression estimators[85] can then be used to make inference.

Simultaneously combining all available evidence

Assume 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 performance of any particular device. To do this requires the assumption that particular relationships among the devices exist although uncertainty

about the relevance of these relationships remains. These uncertainties may relate to: how device 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.

Assume Y_{sjkm} denotes the mth outcome associated with the kth device for the jth group of patients within the sth study. A model that reflects heterogeneity among the outcomes assumes the expected or average outcome, denoted $E(Y_{sjkm})$, can be modeled linearly using a link function g(.), generically as:

$$g(E(Y_{sjkm})) = \alpha_m + \beta_k + \gamma_{mk} + a_s + b_{j(s)} + c_{k(s)} + d_{m(s)}$$

where α_m = average for mth outcome; β_k = the effect of kth 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 sth study; $b_{j(s)}$ = study-specific effect of jth group within sth study; $c_{k(s)}$ = study-specific effect of treatment *k* within sth study; and $d_{m(s)}$ = study-specific effect of outcome *m* within the sth study. The observed outcomes are summaries, for example 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 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 performance of particular devices. Heterogeneity among outcomes is permitted by assuming 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.

The future

Epidemiology and evidence-informed practice and policy

Modern epidemiology, defined as the basic science of public health based on the best available evidence,[86] is strikingly becoming the essential link between an exploding demand for the knowledge derived from diverse evidence and the decisions made in health-care policy and practice settings. In the larger public health context, the imminent future of device epidemiology will be to integrate and infer from massive amounts of heterogeneous and multidimensional data available from disparate data sources. In doing so, 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 the choice of best available treatment targeted to specific populations.

Notably, there is no unified, agreed framework for combining evidence from diverse data sources regarding medical device implants. The strengths and limitations of systematic review, quantitative, interpretive, narrative, sequenced, and other synthesis approaches should be evaluated in the context of specific public health policy and health-care settings toward an accepted framework for evaluation of medical devices.

Epidemiology and regulatory science

The FDA launched, in 2010, the Medical Devices Epidemiology Network (MDEpiNet) Initiative, with the objective to identify evidence gaps and questions, datasets, and approaches for conducting robust analytic studies to improve understanding of clinical outcomes and performance of medical devices through strategic consortium with academic centers. This 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 use of epidemiologic methods and tools can in many instances generate valid, reliable, and objective evidence that is a prerequisite for the application of an evidence-based medicine approach. MDEpiNet structure emphasizes the intersection between meeting regulatory responsibilities and addressing the public health needs surrounding medical devices. An important task for the 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. The intent is to have a comprehensive, up-to-date risk–benefit profile of specific medical devices at any point in its life cycle so that we can make optimally informed decisions and provide more useful information to practitioners, patients, and industry.

Epidemiology and international infrastructure

The accelerating pace of emerging medical technologies will continue, and the information science applications are expected to further shape the information technology -based health care dealing with new demands for storage, transmission, management, and analysis of patient data. The future 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 in the clinical areas where individual countries are limited in developing a research infrastructure. One example of such collaborative effort is the FDA-initiated International Consortium of Orthopedic Registries, led by Cornell University and the Hospital for Special Surgery.[87] An international infrastructure creates opportunities for the development of novel methods for epidemiologic studies. The methods for harmonization, sharing, and combining data are not well developed and require innovative approaches. Furthermore, making conclusions and recommendations when device performance varies from country to country is a unique challenge, often beyond statistics or current paradigms in epidemiology.

References

1. *Espicom. World Medical Market Forecast 2008*. Princeton: Espicom Business Intelligence, 2008.
2. *Health Industries: Medical Device Industry Assessment*. http://www.ita.doc.gov/td/health/medical.html (accessed July 2011).
3. Federal Food, Drug and Cosmetic Act. *Section 201, Definitions (h)*.
4. World Health Organization. *Medical Device Regulations: Global Overview and Guiding Principles*. http://www.who.int/medical_devices/publications/en/MD_Regulations.pdf (accessed Dec 2010).
5. The Council of the European Communities. *Council 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 Dec 2010).
6. Commonwealth of Australia. *Therapeutic Goods Act*. http://www.comlaw.gov.au/ComLaw (accessed Dec 2010).
7. Commonwealth of Australia. *Therapeutic Goods (Medical Devices) Regulations*. http://www.comlaw.gov.au/ComLaw (accessed Dec 2010).
8. Office of the Auditor General of Canada. *1995 May Report of the Auditor General of Canada. Exhibit 4.5—The Definition of a Medical Device under the Food and Drugs Act*. http://www.oag-bvg.gc.ca/internet/English/att_9504xe05_e_4828.html (accessed July 2011).
9. The Global Harmonization Task Force. *Information Document Concerning the Definition of the Term "Medical Device"*, 2005. http://www.ghtf.org/documents/sg1/sg1n29r162005.pdf (accessed Dec 2010).
10. Federal Food, Drug and Cosmetic Act. *Section 513 (a)(1)(A)*.
11. Federal Food, Drug and Cosmetic Act. *Section 513 (a)(1)(B)*.

12. Federal Food, Drug and Cosmetic Act. *Section 513 (a)(1)(C)*.
13. Muni NI, Gross TP, Boam AB, Wang S, Zuckerman BD. Challenges in regulating breakthrough medical devices. *Food Drug Law J* 2005; **60**: 137–42.
14. 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**: 157–79.
15. Hammill S, Phurrough S, Brindis R. The National ICD Registry: now and into the future. *Heart Rhythm* 2006; **3**: 470–3.
16. Ramsay CR, Wallace SA, Garthwaite PH, Monk AF, Russell IT, Grant AM. Assessing the learning curve effect in health technologies. Lessons from the non-clinical literature. *Int J Technol Assess Health Care* 2002; **18**: 1–10.
17. Hannan EL, Racz M, Ryan TJ, McCallister BD, Johnson LW, Arani DT, *et al*. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. *JAMA* 1997; **277**: 892–8.
18. 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**: 2117–27.
19. 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**: 511–20.
20. 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**: 646–54.
21. Nordin P, van der Linden W. Volume of procedures and risk of recurrence after repair of groin hernia: national register study. *BMJ* 2008; **336**: 934–7.
22. Rosenfield KM. Clinical competence statement on carotid stenting: training and credentialing for carotid stenting—multispecialty consensus recommendations. *J Vasc Surg* 2005; **41**: 160–8.
23. 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* **33**: 99–103.
24. 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**: 931–5.
25. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; **342**: 1887–92.
26. Dreyer NA, Garner S. Patient registries, predictive models and optimal care. *JAMA* 2009; **302**: 2662.
27. *American Recovery and Reinvestment Act* (ARRA), Public Law (111–5), 2009.
28. *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC: Institute of Medicine (IOM), 2009.
29. Chalkidou K, Tunis S, Lopert R, Rochaix L, Sawicki PT, Nasser M, Xerri B. Comparative effectiveness research and evidence-based health policy: experience from four countries. *Milbank Q* 2009; **87**: 339–67.
30. 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): S58–67.
31. 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): S121–8.
32. Marinac-Dabic D, Bonhomme M, Loyo-Berrios N, Gross T, Ciperson R, Gardner S, *et al*. Medical devices post-approval studies program: vision, strategies, challenges and opportunities. *Food Drug Law J* 2007; **62**: 597–604.
33. Swedish Hip Arthroplasty Register. http://www.shpr.se/en/default.aspx (accessed July 2011).
34. National Joint Registry of England and Wales. http://www.njrcentre.org.uk/njrcentre/default.aspx (accessed Dec 2010).
35. Australian Orthopedic Association. National Joint Replacement Registry. http://www.dmac.adelaide.edu.au/aoanjrr/ (accessed Dec 2010).
36. Canadian Joint Replacement Registry. http://www.cihi.ca/CIHI-ext-portal/internet/EN/TabbedContent/types+of+care/specialized+services/joint+replacements/cihi021359 (accessed Dec 2010).
37. Bright R. Special applications of pharmacoepidemiology. In: Strom B, Kimmel S, eds. *Textbook of Pharmacoepidemiology*. John Wiley & Sons, Ltd., 2006, pp. 399–445.
38. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 2000; **342**: 1907–9.
39. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin, 2001.
40. Normand SL. Some old and some new statistical tools for outcomes research. *Circulation* 2008; **118**: 872–84.

41. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; **127**: 757–63.
42. 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.
43. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996; **91**: 444–72.
44. Heckman JJ. Comment on 'Identification of causal effects using instrumental variables.' *J Am Stat Assoc* 1996; **91**: 459–62.
45. FDA. *Medwatch Form 3500A*. http://www.fda.gov/downloads/Safety/MedWatch/HowToReport/DownloadForms/UCM082728.pdf (accessed June 2010).
46. FDA. *Manufacturer and User Facility Device Experience Database (MAUDE)*. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/ReportingAdverseEvents/ucm127891.htm (accessed June 2010).
47. Global Harmonization Task Force, Study Group 2. *Post Market Surveillance Final Documents*. http://www.ghtf.org/sg2/sg2-final.html (accessed Dec 2010).
48. Medical device problem reporting for the betterment of healthcare. *Health Devices* 1998; **27**: 277–92.
49. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA Spontaneous Reporting System (with discussion). *American Statistician* 1999; **53**: 177–90.
50. DuMouchel W, Pregibon D. Empirical bayes screening for multi-item associations. *Proceedings KDD 2001*. San Diego, CA: ACM Press, 2001, pp. 67–76.
51. FDA. *Event Problem Codes*. http://www.fda.gov/cdrh/problemcode (accessed June 2010).
52. FDA. *MedSun: Medical Product Safety Network*. http://www.fda.gov/cdrh/medsun/about.html (accessed June 2010).
53. 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, UK: John Wiley & Sons, 2007, pp. 63–77.
54. Samore MH, Evans RS, Lassen A, Gould P, Lloyd J, Gardner RM, *et al*. Surveillance of medical device-related hazards and adverse events in hospitalized patients. *JAMA* 2004; **291**: 325–34.
55. Bright RA, Brown SL. Medical device epidemiology. In: Brown SL, Bright RA, Tavris DR, eds. *Medical Device Epidemiology and Surveillance*. Chichester, UK: John Wiley & Sons, 2007, pp. 22–42.
56. *FDA Public Health Notification: Serious Complications Associated with Transvaginal Placement of Surgical Mesh in Repair of Pelvic Organ Prolapse and Stress Urinary Incontinence*. http://www.fda.gov/medicaldevices/safety/alertsandnotices/publichealthnotifications/ucm061976 (accessed July 2011).
57. *FDA Public Health Notification: Life-threatening Complications Associated with Recombinant Human Bone Morphogenetic Protein in Cervical Spine Fusion*. http://www.fda.gov/medicaldevices/safety/alertsandnotices/publichealthnotifications/ucm062000 (accessed July 2011).
58. *FDA Public Health Notification: MRI-Caused Injuries in Patients with Implanted Neurological Stimulators*. http://www.fda.gov/medicaldevices/safety/alertsandnotices/publichealthnotifications/ucm062125 (accessed July 2011).
59. Reefhuis J, Honein MA, Whitney CG, Chamany S, Mann EA, Biernath KR, *et al*. Risk of bacterial meningitis in children with cochlear implants. *N Engl J Med* 2003; **349**: 435–45.
60. Tavris DR, Dey S, Albrecht-Gallauresi B, Brindis RG, Shaw R, Weintraub W, *et al*. Risk of local adverse events following cardiac catheterization by hemostasis device use—phase II. *J Invasive Cardiol* 2005; **17**: 644–50.
61. Tavris DR, Gallauresi BA, Lin B, Rich SE, Shaw RE, Weintraub WS, *et al*. Risk of local adverse events following cardiac catheterization by hemostasis device use and gender. *J Invasive Cardiol* 2004; **16**: 459–64.
62. Intermacs. http://www.intermacs.org (accessed June 2010).
63. Hefflin BJ, Bright RA. Injuries associated with medical devices: pilot study and surveillance. *The NEISS All Injury Update* 2003; **3**: 4.
64. Wang C, Hefflin B, Cope JU, Gross TP, Ritchie MB, Qi Y, Chu J. Emergency department visits for medical device-associated adverse events among children. *Pediatrics* 2010; **126**: 247–59.
65. *FDA's Sentinel Initiative*. http://www.fda.gov/safety/fdassentinelinitiative/default.htm (accessed July 2011).
66. Mass-DAC. *Percutaneous Coronary Intervention in the Commonwealth of Massachusetts: Fiscal Year 2007 Report*. http://www.massdac.org/sites/default/files/reports/ (accessed July 2011).
67. 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**: 247–56.

68. Resnic FS, Gross TP, Marinac-Dabic D, Loyo-Berrios N, Donnelly S, Normand SL, *et al*. Automated surveillance to detect postprocedure safety signals of approved cardiovascular devices. *JAMA* 2010; **304**: 2019–27.
69. *Postmarket Surveillance. Code of Federal Regulations Title 21*, Pt. 822, 2002 ed. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=822 (accessed Oct 2010).
70. Gliklich RE, Dreyer NA, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ). (Prepared by Outcome DEcIDE Center [Outcome Sciences, Inc. dba Outcome] under Contract No. HHSA290200500035I TO1.) AHRQ Publication Number 07-EHC001-1, 2007.
71. Peterson ED, Kaul P, Kaczmarek RG, Hammill BG, Armstrong PW, Bridges CR, *et al*. From controlled trials to clinical practice: monitoring transmyocardial revascularization use and outcomes. *J Am Coll Cardiol* 2003; **42**: 1611–16.
72. Douglas PS, Brennan JM, Anstrom KJ, Sedrakyan A, Eisenstein EL, Haque G, *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**: 1629–41.
73. Malenka DJ, Kaplan AV, Sharp SM, Wennberg JE. Postmarketing surveillance of medical devices using Medicare claims. *Health Aff (Millwood)* 2005; **24**: 928–37.
74. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005; **58**: 323–37.
75. Astor BC, Kaczmarek RG, Hefflin B, Daley WR. Mortality after aortic valve replacement: results from a nationally representative database. *Ann Thorac Surg* 2000; **70**: 1939–45.
76. Daley WR, Kaczmarek RG. The epidemiology of cardiac pacemakers in the older US population. *J Am Geriatr Soc* 1998; **46**: 1016–19.
77. Greenland S, Finkle WD. A case–control study of prosthetic implants and selected chronic diseases in Medicare claims data. *Ann Epidemiol* 1998; **8**: 319–26.
78. Greenland S, Finkle WD. A retrospective cohort study of implanted medical devices and selected chronic diseases in Medicare claims data. *Ann Epidemiol* 2000; **10**: 205–13.
79. 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**: 1487–90.
80. Bozic KJ, Ong K, Lau E, Kurtz SM, Vail TP, Rubash HE, *et al*. Risk of complication and revision total hip arthroplasty among Medicare patients with different bearing surfaces. *Clin Orthop Relat Res* 2010; **468**: 2357–62.
81. 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**: 526–33.
82. 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**: 47–54.
83. Tchernis R, Horvitz-Lennon M, Normand SL. On the use of discrete choice models for causal inference. *Stat Med* 2005; **24**: 2197–212.
84. Rosenabaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.
85. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med* 2004; **23**: 2937–60.
86. Robson B, Baek OK. *The Engines of Hippocrates: From the Dawn of Medicine to Medical and Pharmaceutical Informatics*. Hoboken, NJ: John Wiley & Sons, Inc., 2009.
87. *Development of Scientific Infrastructure for the Consortium of Existing Orthopedic Registries: Solicitation Number: FDA1077938*. https://www.fbo.gov (accessed July 2011).

CHAPTER 28
Studies of Drug-Induced Birth Defects

Allen A. Mitchell
Slone Epidemiology Center at Boston University, Boston University Schools of Public Health and Medicine, Boston, MA, USA

Introduction

Teratogenesis is a very different phenomenon from other drug-induced hazards, and it therefore requires special consideration. Although the fetus may experience a wide range of adverse effects as a result of antenatal drug exposure, such as mental/ motor deficits and learning and behavioral problems, for reasons that include their rarity, often dramatic presentation, and their association with thalidomide, this chapter will confine itself to issues surrounding drug-induced physical malformations.

The nature of birth defects and their relation to drugs

Birth defects are part of the human condition, having been observed throughout history. Major birth defects, typically defined as those that are life threatening, require major surgery, or present a significant disability, affect approximately 3–4% of liveborn infants.[1,2] Minor malformations are of lesser clinical importance, and estimates of their prevalence vary considerably because of substantial differences in definition and detection.

Over the centuries, the "deformities" that characterize most birth defects frequently have been viewed as a punishment to the mother or family for some fault or transgression on their part. This view was undoubtedly reinforced by the rarity of birth defects, their unpredictable occurrence, and the absence of known causes. Perhaps because these factors have not changed very much over time, elements of this primitive view persist today, largely in the form of guilt. Parents tend to search their memories to identify some factor—any factor—that might account for their misfortune. In developed societies, attention often focuses on drugs taken in pregnancy.

This concern, of course, is not without foundation. While it was believed that the placenta protected the fetus from noxious agents as recently as 70 years ago, that belief was shattered in 1941 by the recognition that maternal rubella infection in pregnancy produced a distinctive pattern of birth defects among exposed infants.[3] Two decades later, the thalidomide disaster demonstrated that drugs, too, could be teratogenic.[4] Many thousands of infants were born with major limb reductions and other defects, and the tragedy of this epidemic was etched into the consciousness of medical practitioners and the public alike. In the years that followed, other drugs were shown to be teratogenic, ranging from valproic acid to isotretinoin. Additional drugs were alleged to be teratogenic, and although many of those allegations were unsupported by subsequent studies, they served to reinforce the general concern about the teratogenic effects of marketed drugs.

Teratogenesis is a unique kind of adverse drug effect, since it affects an organism (the fetus) other than the one for whom the drug was intended (the

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

mother). In a benefit/ risk consideration, the fetus may at best indirectly benefit from a medication given to its mother (e.g., by an improvement in the mother's health), but the fetus alone is at risk for birth defects. That "innocent bystander" status of the fetus raises profound medical, moral, and legal issues. It also poses serious concerns about the consequences of allegations that a given drug may be teratogenic.

Can teratogenesis be predicted prior to marketing a drug?

Under ideal circumstances, one would identify the teratogenic potential of drugs before they were used in humans. Unfortunately, our ignorance about the basic mechanisms of organ formation has constrained development of predictive *in vitro* tests. Testing in animals may be helpful in certain circumstances; for example, vitamin A and its congeners produce consistent patterns of malformations across many different species. However, for most known human teratogens, results of animal tests vary so much as to seriously limit their predictive value.[5] Further, understanding of the structure/ activity relationships of a particular agent can help predict that drug's efficacy and some adverse effects, but it does not necessarily help predict its teratogenic potential. Because there are no theoretical, *in vitro*, or animal models that can reliably and consistently provide meaningful information about the likelihood of human fetal risk, we are usually completely unaware of a given drug's teratogenic potential when it is first used in humans.

Clinical premarketing studies cannot be expected to provide this information either. Information derived from experience among non-pregnant adult women is not informative when the concern is teratogenesis, and might even provide a false sense of reassurance—it is worth recalling that thalidomide was used as a sedative in pregnant women specifically because of its "safety profile" in non-pregnant adults. Traditionally, women of childbearing age were excluded from early clinical studies, specifically because of concerns about potential teratogenicity. Newer guidelines are designed to reverse many of these exclusions; pregnant women will be enrolled in rare instances where a drug is indicated in pregnancy, but the numbers targeted to assess benefit are insufficient to identify risks of specific birth defects. More typically, trials that may now include women of childbearing potential will take pains to assure that study subjects are at minimal risk of becoming pregnant.

A note about over-the-counter drugs

Over-the-counter (OTC) drugs present a unique situation. Whatever caution physicians might exercise in their prescribing of drugs to pregnant women, they have little control over what consumers purchase OTC. It can reasonably be assumed that women of childbearing age, like other consumers and their physicians, consider OTC drugs to be safer than prescription products, and they may assume the same is true for use of these drugs in pregnancy. While prescription drugs tend to become available OTC on the basis of a history of wide use and safety in the general population, the process of switching to OTC availability rarely takes into account a drug's risk or safety with respect to the fetus (usually because such data are unavailable). This is particularly the case for drugs that became available OTC decades ago. As noted below, teratogenicity is more difficult to assess when a drug is used without prescription, and it is ironic that we may know less about the teratogenic hazard of drugs available OTC than we do about prescription drugs.

The important reality is that we lack an understanding of the teratogenic effects of newly marketed and most established prescription drugs as well as most OTC drugs. The opportunity to gain such knowledge comes from postmarketing experience, where pharmacoepidemiology can and must play a crucial role.

Clinical problems to be addressed by pharmacoepidemiologic research

Like other adverse drug effects, teratogenesis is a critical aspect of a drug's benefit/ risk profile, and such information obviously should be available to prescribers and consumers. Unlike other adverse

drug effects, however, teratogenesis raises uniquely important and controversial clinical issues. First, the fetus is the "innocent bystander" with respect to its mother's therapy. Second, teratogenesis is not a concern limited to women who are pregnant when drug treatment is initiated; since roughly half of pregnancies (at least in the US) are unplanned, teratogenesis must also be a concern among women who might become pregnant while taking a medication. Finally, unlike other adverse outcomes, teratogenic effects can be prevented by avoidance of pregnancy, and the birth of a malformed infant can be avoided by termination of pregnancy. Our understanding of a drug's teratogenic risk therefore has important consequences for how a given drug is used clinically, and options the treated patient might elect.

Drugs known to be teratogenic

Experience has shown that human teratogens tend to fall into two broad categories.[6] Drugs that produce major defects in a high proportion (roughly 25%) of exposed pregnancies can be considered "high-risk" teratogens (e.g., thalidomide and isotretinoin). More common are "moderate risk" teratogens, which increase the rate of specific birth defects by perhaps 5- to 20-fold (e.g., carbamazepine and neural tube defects). In the latter situation, for example, the background rate of neural tube defects of 1 in 1000 pregnancies might be increased to 10 in 1000. The differences between high-risk and moderate-risk teratogens have relevance for how these drugs are considered in the clinical setting.

Generally speaking, there are three approaches applied to the few drugs known to be human teratogens. In rare instances, such as was the case for thalidomide in most countries, a drug may be prohibited or removed from the general market once its teratogenicity becomes known; for thalidomide in the 1960s, this approach was justified by the fact that this high-risk teratogen posed a large absolute risk to the fetus but did not offer important or unique therapeutic benefits to the women using the drug.

Most known teratogens, such as phenytoin and valproic acid, pose moderate risks and are often considered to fill an important clinical need. For such drugs, information about teratogenicity is provided to physicians, who are expected to discuss the benefits and risks with their patients, who in turn can make informed decisions about their drug treatment. In some settings, a drug may be restricted to prescription by selected physicians (see thalidomide example below); however, until follow-up data are available, the effectiveness of this approach remains unclear. It should be noted that considerations in these doctor–patient discussions may differ according to each woman's risk of becoming pregnant while on the drug.

The third approach utilizes a formal risk management program designed to minimize the risk of exposures during pregnancy (see Chapter 29). Such programs may involve education of physicians and patients that is combined, in some cases, with restricted access to the drug. The educational component is intended to assure that physicians and their patients are informed about the drug's teratogenicity and the importance of avoiding pregnancy. The first such effort, in the US, was initiated in late 1988 by the manufacturer of isotretinoin (Accutane®), a high-risk teratogen that is uniquely effective in the treatment of severe acne.[7] Preliminary data from this voluntary "pregnancy prevention program" suggested some success in achieving both educational objectives, reducing the number of exposed pregnancies from about 4 per 1000 courses of therapy in 1989 to about 1 per 1000 in 2002.[8,9] While more restrictive approaches have since been applied to Accutane and its generic equivalents ,[10,11] it is as yet unclear whether they have further reduced the rate of pregnancy among isotretinoin-exposed women.

In July 1998, newly identified uses for thalidomide prompted the US Food and Drug Administration to approve, for the first time in the US, marketing of the drug (as Thalomid,® to prevent skin symptoms of erythema nodosum leprosum), but only with an unprecedented FDA-regulated program sponsored by the drug's manufacturer, Celgene Corporation. This program included an educational component similar to that used for Accutane®, but also restricted prescription and distribution of the drug to registered prescribers and pharmacies,

respectively; it also mandated that all patients participate in a follow-up survey designed to monitor and enhance compliance with the manufacturer's "System for Thalidomide Education and Prescribing Safety (STEPS)." Experiences derived from that effort led to a revision in the program in 2001, in which patients and prescribers complete a telephone-based screening before initial therapy and with each subsequent prescription. This screening provides an authorization number to the prescriber; that number, placed on the prescription and verified by the pharmacist, is designed to assure compliance with the risk management program. While this approach is more rigorous than those that were used for isotretinoin or other teratogens, it is important to recognize that most women of childbearing age who receive thalidomide have multiple myeloma and other resistant forms of cancer, and it remains an open question whether the low rates of pregnancy encountered in this risk management program can be projected to other populations.

It appears that no single formula for pregnancy prevention will apply to all human teratogens, but the efforts focused on isotretinoin and thalidomide will, with appropriate, rigorous, and independent evaluation, advance our understanding of how best to balance the therapeutic benefits of known human teratogens against the risks of fetal exposure.

Drugs for which teratogenic risk is unknown

For reasons described above, the vast majority of prescription drugs and virtually all non-prescription drugs fall into the category of drugs for which teratogenic risk is unknown. Their labels may include a general warning against unnecessary use in pregnancy or to consult a physician (who cannot provide meaningful advice in the absence of data!), but such cautions hardly contribute to rational drug therapy. In settings where the true teratogenic risk is nil, these warnings discourage potentially useful therapy; where the true risk is elevated (but unidentified) for a particular drug, the "standard" warning offers little practical discouragement to its use in pregnancy.

Drugs for which teratogenesis is alleged . . . and clinical consequences

At one time or another, a large number of drugs have been alleged to be teratogenic, and the clinical consequences can be profound. In one notorious situation, allegations of teratogenicity resulted in a widely used drug being withdrawn from the market. In the late 1970s and early 1980s, the antinausea drug Bendectin® (Debendox®, Lenotan®), used widely to treat nausea and vomiting of pregnancy, was alleged to cause a variety of birth defects; the history of this experience has been reviewed.[12] Ironically, the aggregate data on the teratogenic hazards of Bendectin® have ultimately provided the strongest evidence of safety for any medication used in pregnancy. Despite that evidence, however, the manufacturer withdrew the drug from the market because of active and potential litigation. At least one study suggests that hospital admissions for hyperemesis gravidarum increased significantly following the drug's withdrawal,[13] and there is concern that, in the absence of Bendectin®, women will be treated with other antinausea drugs (e.g., odansetron), for which the teratogenic risks are largely unknown.

Other clinical aspects of unproven allegations are given too little attention. Upon learning that a drug she took in pregnancy might be teratogenic, a woman who has given birth to a malformed infant may become overwhelmed with feelings of guilt. Further, an exposed woman who is currently pregnant may develop considerable anxiety that can lead to a number of clinical consequences, ranging from medical consultations, to diagnostic procedures (e.g., amniocentesis), to elective termination of the pregnancy. An experience in the mid-1970s involving a non-drug exposure is instructive. Following widely publicized allegations that spray glue adhesives were teratogenic, a US regulatory agency withdrew the product from the market. Although the allegations were subsequently found to be without basis, a survey of genetic counseling centers identified 1100 inquiries from pregnant women prompted by concern about exposure to these agents; moreover, 11 underwent amniocenteses and because of this exposure, nine women had therapeutic abortions (one had vague evidence

of chromosomal damage and eight had no evidence of malformation).[14] These unique and potentially serious clinical consequences of false-positive allegations argue for heightened attention to scientific rigor and caution in the teratologic assessment of drugs.

The fallacy of "class action" teratogenesis

Another clinically important concern specific to teratogenesis is the issue of "class action." It is widely recognized that an understanding of structure/activity relationships shared by members of a given drug class can be helpful in predicting a given class member's efficacy and adversity (indeed, this view is incorporated into regulatory action in the form of class labeling). However, class-based pharmacologic effects cannot be assumed to hold when the adversity at issue is teratogenesis. Given our ignorance about the causes of most birth defects, we cannot know whether it is the chemical structure common to the class that is responsible for teratogenesis or whether the responsible component is that part of the structure that differentiates one class member from another. For example, thalidomide and glutethimide (Doriden® and other brands) are both glutarimides, and both are sedative/ hypnotics. Despite their structural and clinical similarities, thalidomide is clearly a high-risk teratogen and glutethimide is not.[15] 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 reassurance about the safety of one drug can be extended to other members of that drug's class.

Methodologic problems to be addressed by pharmacoepidemiologic research

In many ways, the epidemiologic issues involved in the study of birth defects are similar to those of other adverse outcomes; these are considered in detail elsewhere in this text (see Chapters 3 and 4, and Part V). However, there are a number of considerations that are unique to birth defects or are sufficiently important to warrant particular attention. These have to do with sample size considerations, definitions of exposure and outcome, confounding, and biologic plausibility.

Although serious birth defects occur in approximately 3–4% of liveborn infants, we cannot consider "birth defects" as a single, homogeneous outcome. In fact, physical birth defects include a wide range of malformations that vary in many ways, including their gestational timing, embryologic tissue of origin, and mechanism of development. As examples of variations in timing of occurrence, chromosomal abnormalities generally predate conception; neural tube defects develop in the earliest weeks of gestation; cleft palate develops later in the first trimester; and microcephaly can develop relatively late in pregnancy. As an illustration of variations in embryologic tissue of origin, some cardiovascular malformations (but not others) are derived from the neural crest cells that migrate from the area surrounding the primitive neural tube. Variations in mechanisms of development include inhibition, disruption, or alteration of the embryologic tissue that is responsible for normal structural development. From a theoretical perspective, then, one would predict that the malformations produced by a drug would vary according to the timing of exposure, the sensitivity of the end organ (i.e., embryologic tissue), and the mechanism of its teratogenesis.

Experience supports what would be predicted from biology. Even a brief review of known teratogens reveals a fact that is highly relevant in pharmacoepidemiologic studies: teratogens do not uniformly increase the rates of all birth defects, but rather increase rates of selected defects. Thus, the "classic" high-risk teratogen thalidomide produces defects in about 25% of exposed infants, but that overall increase is largely the result of increases in limb, spine, and central nervous system malformations;[16] another high-risk teratogen, isotretinoin, affects a similar proportion of liveborn infants, but again, that overall rate is the result of increases in rates of selected specific defects (ear, central nervous system, and cardiac).[17] Moderate-risk teratogens also increase the rate of specific defects (though to a lesser degree): valproic acid most notably increases

rates of neural tube defects,[18] warfarin increases the rate of cartilage defects,[19] and ACE inhibitors increase rates of renal defects.[20]

Sample size considerations

The fact that pharmacoepidemiologic studies must consider rates of *specific* birth defects has a dramatic effect on sample size requirements, both for estimating risk and providing assurances of safety. With respect to risk in a population of women taking a given drug, a cohort study with a sample size of a few hundred exposed pregnancies might be sufficient to identify a doubling of the 3–4% overall rate of birth defects; ruling out a doubling of the overall rate would require larger numbers, but these would still be within the same order of magnitude. However, each specific defect occurs with far less frequency, ranging from about 1 per 1000 live births for oral clefts to 1 or fewer per 10 000 for biliary atresia. For a cohort study to *detect* a doubling of risk for a relatively common specific birth defect (e.g., 1/1000) requires a sample size of over 20 000 exposed pregnancies (see Chapter 4 and Appendix A). To *rule out* a doubling of risk for the same defect, one would need a far larger sample size of exposed pregnancies.

Exposure—and exposure misclassification

There are aspects of random exposure misclassification that require special consideration in birth defects studies. One relates to medications which, though taken, are not identified as exposures, and another relates to medications identified as exposures that were not taken.

Medications taken but not identified as exposures

Most pharmacoepidemiologic research focuses attention on prescribed drugs not only because (as noted above) OTC drugs are widely perceived as "safe" but also because many studies utilize data sources with missing or inadequate information on non-prescribed drugs. As a category, non-prescribed drugs also includes herbal, vitamin/ mineral, and supplement products, and their effects on the fetus are also largely unstudied. Ironically, for decades these agents have comprised the majority of drug exposures in pregnancy.[21–23] Use of herbal products by pregnant women[24,25] raises additional concerns, given the almost unregulated nature of these agents and the fact that the precise contents and purity of these products are often unknown. It is therefore important that pharmacoepidemiologic research include consideration of the wide range of non-prescribed drugs.

In this context, "non-prescribed" can include prescription drug products that were "borrowed" from friends, neighbors, and relatives: more than 25 years ago, we found that over 20% of exposures to selected prescription drugs (e.g., Darvon [propoxyphene]) can come from such "non-prescribed" sources.[26] Others have since documented that about 25% of adults [27] and 20.1% of teenage girls[28] reported borrowing or sharing prescription medications, and nationally representative US data reflected that 36.5% of women of reproductive age acknowledged borrowing or sharing prescription medications.[29] Such exposures, of course, are unlikely to appear in a pregnant woman's medical record. For prescription drugs, one might question the validity of drug exposure information derived from records (medical, billing, insurance, etc.) versus that derived from patient interviews (see Chapter 41), but there is little question that information on the use of OTC, herbal, or borrowed prescription drugs must be obtained directly from the patient.

Illicit drugs represent a distinct but important subset of non-prescribed drugs. Use of these drugs is seriously under-reported, whether information is drawn from records or interviews. Except in rare settings where exposure may be identified through systematic screening of biologic samples (e.g., urine, blood), epidemiologic studies have major limitations in identifying teratogenic (and confounding) effects of illicit drug use in pregnancy.

Medications identified as exposures but not taken

While the medical record may not capture exposures to various drugs, including borrowed prescription drugs, prescription medications that do

appear in the medical record may not reflect actual exposure; first, women may not fill prescriptions issued—a recent study of electronic prescription records in Massachusetts found that 28% of almost 200 000 prescriptions written were not filled.[30] This problem of primary non-adherence is compounded by the widespread but difficult-to-document problem of secondary non-adherence—that is the failure of patients to take the medications as recorded on the prescription (see Chapter 42).

Gestational timing of exposures

As noted above, various organs form at different times in gestation, and for that reason, timing of exposure takes on particular importance. An exposure cannot cause a birth defect involving a given organ after that organ's formation is complete, and, in general, exposures that occur well before development of a given organ are unlikely to cause malformations of that organ. Random misclassification of exposures with respect to gestational timing reduces the likelihood of detecting teratogens, so timing of drug use takes on particular importance in birth defects research. Since data sources vary in the accuracy of information on pregnancy onset, they also will vary in the accuracy of gestational timing.[31]

Recall bias

Because of the sample size requirements described above, many researchers have turned to the case–control approach for the study of specific birth defects. Such studies generally (though not always) rely on maternal interviews for exposure information, and this approach raises concern both about the overall accuracy of recall (see also Chapters 3 and 41) and its susceptibility to bias. More than other drug-induced adverse outcomes, the birth of a malformed child carries an emotional burden and guilt that may affect recall of exposures in pregnancy. When compared to a mother of a normal child, the mother of a malformed infant may be more likely to recall carefully every possible act, event, and drug exposure in pregnancy.[32] This tendency is reinforced by repeated inquiries from physicians, nurses, genetic counselors, and relatives, as well as by media and legal attention on the subject of drug-induced birth defects. Thus, in a setting where drug exposure is in fact identical among mothers of normal and malformed infants, one might predict that recall of exposure will be more complete among the latter than among the former, creating a false association between the drug and the birth defect. Concern about recall bias is more than theoretical—such bias may well explain a number of drug–defect associations[33,34] that have subsequently been refuted.[35,36] On the other hand, evidence supporting the role of recall bias is inconsistent, and the issue of when and to what extent recall bias is present remains an unresolved controversy (see Chapter 41).

Despite this concern, the simple possibility of recall bias does not invalidate interview-based studies, and there are a number of approaches to reducing and dealing with this problem. These include the choice of controls, the design of the questions, and direct attempts to identify potentially biased recall.

There are differing schools of thought regarding what constitutes an appropriate control for a malformed infant—should controls be infants without malformations, or should they be infants with malformations other than the one under study? Some argue that normal infants should be used because of the possibility that a drug might increase the risk of all malformations, a finding that would be missed if only malformed controls were used.[37] Others argue that, since no known teratogen uniformly increases the risk of all malformations, normal controls may be unnecessary; further, use of normal controls might increase the opportunity for differential recall of exposure between mothers of normal and malformed infants. They argue that controls should comprise infants with a wide range of malformations other than the ones in the cases.[38] By assuring that the controls include a wide range of malformations, one reduces the likelihood that the controls will be biased by inclusion of a large proportion of defects that might be associated with the exposure under study. By restricting comparisons of exposures to those reported by mothers of malformed infants (whether cases or controls), one limits the likelihood of recall bias.

Both approaches are imperfect. Although no teratogen has yet been identified that uniformly increases the risk of all specific defects, the history of teratology is replete with examples of assumptions that proved to be false. We used to think that the fetus was protected by the placenta from noxious agents, and only 30 years ago it was inconceivable to some that a drug (diethylstilbestrol) could produce cancers in the adult offspring of exposed mothers, or birth defects in the children of women exposed to the drug *in utero*. In an effort to avoid such hubris, some researchers,[38] including ourselves, have elected to use two control series, one of malformed and one of normal infants. Since we believe that concern about recall bias exceeds concern about failing to identify an "across-the-board" teratogen, we usually give primary consideration to findings derived from comparisons with malformed controls.

By definition, recall bias cannot exist if reporting of drug exposure is complete among cases and controls. The closer one comes to that ideal, the less the likelihood of recall bias. It thus becomes critical how one elicits exposure information. Studies that use open-ended questions about drug exposure invite differential recall between mothers of malformed and normal infants.[33] As might be predicted, the more specifically one asks questions about drug use, the more likely one is to obtain complete information (see also Chapter 41). Recall is also substantially increased when women are asked about use according to various indications, and it is further increased when drugs are asked by specific names[26] (see also Chapter 41). This approach is not likely to result in exaggerated recall (i.e., false positives), as demonstrated by the fact that use of a specific drug ascertained by such a questionnaire was the same as that estimated from the manufacturer's marketing data; in addition, women do not tend to report exposure to a nonexistent drug.[35] In short, by improving ascertainment of drug exposure among both cases and controls, a carefully designed questionnaire can substantially reduce the opportunity for recall bias.

Unfortunately, the possibility of recall bias cannot be eliminated completely, either by the use of a malformed control group or by asking specific questions about drug use. In an effort to identify women who might be most at risk for biased recall, we began, in 1976, to ask routinely whether a woman has heard that *any* drug affects the risk of any defect. (This series of questions is asked at the end of the interview, so as not to itself affect reporting of exposures and events.) Our *a priori* assumption is that a woman who acknowledges that a particular drug causes (or prevents) a particular defect is more at risk for differential recall than one who does not. This approach has enabled us to identify indirect evidence of biased recall: in our study of the possible protective effects of folic acid on the development of neural tube defects, we observed different risk estimates when we stratified subjects according to their knowledge of the hypothesis.[39] By simply asking women about their perceptions of the teratogenic effects of drugs, one might obtain insight into the nature of biased recall in the study population.

Outcome

Given the etiologic heterogeneity of malformations, some have attempted to classify birth defects according to specific categories. We were among those who, more than two decades ago, classified defects by organ system, such as "musculoskeletal" or "cardiovascular."[1] However, classification in this way has little embryologic or teratologic basis, and a more appropriate approach is to create categories that reflect the embryologic tissue of origin. For example, neural crest cells in the earliest stages of embryogenesis migrate to form a variety of structures, including those of the face/ears, parts of the heart, and the neural tube.[40] Interference with the normal development of the neural crest would therefore be expected to produce malformations of tissues derived from neural crest, and that phenomenon has been observed in a number of animal experiments. In fact, these patterns have also been observed for certain human teratogens, the most striking example of which is the retinoid isotretinoin, which interferes with neural crest cell migration/ development and leads to specific malformations of the ear, heart, and neural tube.[17]

Similarly, certain defects are believed to result from disruption of the embryonic vasculature.[41,42] Although our ignorance about the origins of most birth defects may limit our ability to create categories that share a common etiology, it is preferable, whenever possible, to classify birth defects according to an understanding of their embryologic origins.[43]

Confounding

As with any other aspect of pharmacoepidemiologic research (see Chapters 3 and 47), confounding must be taken into account in studies focused on birth defects. Among those variables that require routine consideration are maternal age, race, geography, and socioeconomic status. An understanding of the epidemiology of a given defect or exposure often identifies other variables which may act as confounders in a specific analysis. For example, ethnic background is strongly related to the risk of neural tube defects, maternal age is a strong risk factor for gastroschisis, and alcohol consumption has been associated with defects derived from neural crest. Since medication use may be associated with various other health behaviors (e.g., vitamin use is more common among nonsmokers than smokers), one may need to consider health behaviors, including nutrition, in studies of certain exposures and outcomes. Further, it may be critically important to separate the teratogenic risk of a drug from the underlying risk associated with the condition for which the drug is taken, something called "confounding by indication"[44] (see Chapters 3, 37, and 47).

Finally, an issue unique to the epidemiologic study of birth defects is the possibility of pregnancy termination. As more malformations become detectable at earlier stages of pregnancy (and as more such pregnancies are terminated), studies of liveborn and stillborn infants will increasingly underestimate the prevalence of such defects. In addition, there are a number of instances where terminations (whether spontaneous or induced) must be considered as a potential confounder (e.g., periconceptional vitamin exposure and neural tube defects).

Biologic plausibility

Our ignorance about the biologic mechanisms by which most human birth defects occur complicates our ability to determine when a finding may be biologically plausible. 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 defects derived from neural crest cells among infants exposed to retinoids,[17] the decreased risk of neural tube defects among infants exposed to folic acid[45] and the increased risks of those defects among infants exposed to folic acid antagonists,[46] and the increased risk of defects resulting from vascular disruption among infants exposed to aspirin and possibly pseudoephedrine.[47,48] However, biologic mechanisms remain unknown for most well-accepted drug–defect associations.

In light of this inconsistency, how does one evaluate the importance of biologic plausibility in relation to newly observed associations? On the one hand, a requirement that every association have an identifiable biologic mechanism would have led to dismissal of virtually every accepted human teratogen. On the other hand, some aspects of biologic plausibility must be met. For example, it is implausible that a defect could be caused by an exposure if that exposure first occurs after the gestational development of the defect has been completed. While less absolute, it is unlikely that an exposure would produce a range of defects which span gestational timing from preconception to late pregnancy and which do not share embryologic tissue of origin. Thus, we cannot dismiss hypotheses simply because they lack a biologically plausible explanation; however, until they are supported by subsequent studies, such hypotheses must be considered more speculative than those hypotheses for which there is a strong biologic basis.

Currently available solutions

There are a variety of approaches that are used to generate and test hypotheses regarding drugs and birth defects. The purpose of this section is not to

list every available design or data set, but rather to describe the types of resources and their respective strengths and weaknesses. For convenience, these may be divided into cohort and case–control designs. Approaches that involve the monitoring of birth defects without the systematic collection of exposure information are not directly applicable to pharmacoepidemiologic study, and are not considered in this chapter; interested readers are referred to an excellent review.[49]

Cohorts

Broadly speaking, there are three types of cohorts relevant to the pharmacoepidemiologic study of birth defects. These are studies designed to follow large populations exposed to various agents, the use of data sets created for other purposes, and follow-up studies of selected exposures.

Studies designed to follow large populations exposed to various agents

This approach involves the identification of a population of pregnant women to be followed, with periodic collection of information on demographic characteristics, exposures, and potential confounders, as well as formal evaluation of the offspring at birth and perhaps at some years later. A number of studies of this kind have been conducted in various countries.[50–53] An example is the US Collaborative Perinatal Project (CPP), which enrolled over 58 000 women between 1959 and 1965, obtained detailed information on their pregnancies, and followed the children until age 7.[1] More recently, researchers in Denmark have assembled a cohort of 100 000 pregnancies,[54] and, in the US, the National Children's Study [55] has initiated a similar effort. The strength of this type of approach lies in the prospective, systematic, and repeated collection of information that includes exposure to a wide variety of medications taken by a diverse population, many potential confounding variables, and good outcome information.

In the CPP, there was sufficient power for commonly used drugs (such as aspirin in the 1960s) to assess risks for malformations overall as well as for certain subgroups of malformations.[56] However, despite the large number of pregnancies in the database, a major weakness of even cohorts in the tens of thousands is the small sample sizes of infants with *specific* malformations. For example, there were approximately 2200 infants with any major malformation in over 50 000 pregnancies followed by the CPP. Among these, there were only 31 with cleft palate (CP) and 11 with tracheoesophageal fistula (TEF). This weakness is further compounded by limited numbers of women exposed to most drugs. For a commonly used drug, taken by as many as 10% of the women, the expected number of exposed infants with CP and TEF would be three and one, respectively; if a drug were used by 3% of pregnant women, the expected numbers would be one and 0.3. Such cohorts may be large enough to identify some high-risk teratogens; however, based on experience with the CPP, even cohorts twice the size are likely to have inadequate power to identify moderate-risk teratogens among commonly used drugs, and power is routinely inadequate to identify such teratogens among the vast majority of other drugs. Further, these large cohorts typically limit enrollment and data collection to a few years; yet, because patterns of drug use change over time, the clinical relevance of the available data becomes limited.

Use of data sets created for other purposes

In recent years, researchers have focused increasing attention on cohorts identified from databases produced for purposes other than for epidemiologic research by organizations or governments involved in medical care (see Part IIIB). The strengths and weaknesses vary with the nature of the specific data set. All have the advantage of identifying exposures independent of knowledge of the outcome, some may have good reporting of malformations, and some may be derived from large populations. Like most other cohorts, studies based on data from health maintenance organizations (HMOs) or insurance plans may be limited by their small samples of specific malformations. For example, a study from Group Health Cooperative in Seattle, Washington, identified almost 7000 pregnancies in which 33% of women filled a prescription for Bendectin®. Among these women, there were a total of 80 malformations identified,

of which only 24 were exposed to the drug *in utero*.[57] Typically, a given exposure is far less prevalent, and sample size constraints are even more striking. [58]

Record linkage systems, such as those from Scandinavia, and particularly Denmark,[59,60] offer promise of better information on exposure and outcome variables. Like other automated data sets, however, they lack important information on potential confounding variables and other factors of potential importance (see below).

In the US, the Sentinel Network represents an ambitious effort underway to take advantage of multiple sources of electronic medical records (See Chapter 30), and an amalgamation of data from a US HMO network has been initiated specifically to consider drug risks in pregnancy (called MEPREP).[61] However, both Sentinel and MEDPREP have yet to demonstrate their validity and utility as tools to generate and test hypotheses. Besides the general challenges of capturing accurate information on exposure, outcomes, and confounding variables, the study of birth defects poses additional challenges, among which are the need to effectively link maternal and infant records, identify gestational timing of exposure with relative accuracy and, as noted above, to include information on potential confounders such as diet, alcohol, smoking, and use of OTC folate and multivitamins at the time of conception. They are also unlikely to identify exposure to non-prescription drugs, which may be of importance both as potential confounders and as primary exposures.

Follow-up of selected exposures

Various mechanisms may be used to identify cohorts of women exposed to specific drugs. Pregnant women can be enrolled in registries by physicians or by the women themselves, often on the basis of a call to a teratogen information service. The strength of such approaches lies in the ability to identify women exposed to a drug of interest early in pregnancy and, most importantly, to identify and enroll the woman before the pregnancy outcome is known. This design offers the additional advantage of providing an opportunity to prospectively collect other information, such as data relating to other exposures and potential confounding variables. These follow-up studies, often called pregnancy registries, have had strong support in the US from the Food and Drug Administration, and sponsors of selected new prescription drugs have been encouraged to establish such registries for those products.[62] The potential value of these cohorts is reflected in the dramatic observation among only 36 women who were followed after first-trimester exposure to isotretinoin:[17] there were 28 liveborn infants, and five (18%) were malformed. More striking than the overall rate of malformation was the distribution of defects: each of the five affected infants had at least one of the specific malformations hypothesized (from premarketing animal studies) to result from isotretinoin exposure (ear, palate, chin, certain heart, and certain brain defects). These findings have been supported by a subsequent study of 94 prospectively identified exposed pregnancies that resulted in live births.[63]

Cohorts of a few dozen to a few hundred exposed pregnancies are highly efficient and effective for identifying—and for ruling out—high-risk teratogens.[64] However, such cohorts are quite limited in their ability to identify a drug as a moderate-risk teratogen or to rule out such an effect. For example, researchers identified 174 women who sought counseling because of first trimester exposure to fluoxetine. While that study provided relatively early evidence of neonatal complications among the offspring of women with *in utero* exposure to fluoxetine, there were few infants with any kind of major malformation.[65] For a specific defect (e.g., oral cleft) with a baseline rate of 1 in 1000 births, such a sample would easily miss increases in risk in the order of 30-fold. On the other hand, even small numbers of exposed/ malformed may have value; for example, that report identified two exposed cases of a rare outcome (persistent pulmonary hypertension of the newborn [PPHN]); this signal was later supported by a large case–control study of PPHN.[66]

Pregnancy registries may be limited by problems of self-referral bias and losses to follow-up (with biases introduced if response to follow-up is related to whether the infant is malformed or normal). In

addition, there may be difficulties in making comparisons between the drug-exposed cohort and those "unexposed." Some registry investigators compare observed rates of defects to rates reported in various other populations, and others compare the observed rates to those among pregnancies with exposures to no drugs or to other drugs (e.g., presumed "non-teratogenic" drugs); both are imperfect, and comparisons rarely consider confounding by indication—the risk of birth defects due to the condition for which the drug in question was taken.[44]

These methodologic concerns aside, all cohort studies have a common limitation: as is clear from the earlier discussion, cohorts are well suited to identify and rule out high-risk teratogens and may occasionally identify meaningful "signals" related to specific birth defects, but it is extremely difficult for such studies to achieve sample sizes sufficiently large to provide the broader range of critical information about a drug's risk or safety; such inquiry requires information not only on the risk of birth defects overall, but must also include sufficiently large samples to permit study of risks of specific defects, most of which have background rates ranging from 1 per 1000 to 1 per 10 000. Thus, while a cohort of 100 or even 1000 exposed pregnancies might provide reassurance that the drug is not another thalidomide or isotretinoin, such a cohort cannot assure that a drug is relatively safe with respect to oral clefts, gastroschisis, or other specific birth defects.

Case–control studies

The rarity of birth defects in general, and of specific defects in particular, argues for the use of the case–control design in pharmacoepidemiologic studies of birth defects when there is a high enough prevalence of exposure to the drug(s) of interest. Of course, the strengths and limitations of these studies are similar to those for case–control studies of other outcomes (see Chapter 3), and will not be reviewed here. Such studies may be conducted on an *ad hoc* basis or within the context of case–control surveillance (see Chapter 19). Examples of the latter are few; in North America, two current examples include the longstanding "Birth Defects Study" conducted by our own group,[35] and the more recently established National Birth Defects Prevention Study, involving a number of state birth defects surveillance programs and coordinated by the US Centers for Disease Control and Prevention.[67]

From the unique perspective of birth defects, case–control studies can have the statistical power required for the assessment of both risk and safety. At the same time, however, they have the potential limitation of biased recall. There are numerous examples that illustrate both issues, but the following is among the most instructive.

In a study of spermicidal contraceptives, researchers using data from an HMO found the prevalence of birth defects among 763 infants born to exposed mothers to be 2.2% ($n = 17$), whereas the rate among infants born to non-exposed mothers was 1.0%. The excess was attributed to four different defects: chromosomal defects, limb reduction defects, hypospadias, and neoplasms.[68] Other investigators used different cohorts to test the hypothesis. Although two analyses involving populations of about 35 000 and 50 000 pregnant women failed to confirm an overall increase in malformation risk,[69,70] neither study had sufficient power to rule out, with reasonable confidence, an increased risk for each of the specific defects identified in the first study. Therefore, two case–control studies were mounted specifically to test the hypothesis. One identified about 100 to 400 cases of each of the outcomes of interest, and for a variety of exposure duration intervals found odds ratios close to unity; more importantly, the upper 95% confidence intervals were 2.2 or lower.[71] The other identified 151 fetuses with trisomy, including 92 with trisomy 21; point estimates for various exposure duration intervals approximated unity, and the study had sufficient power to rule out more than a twofold increase in the risk of trisomy in relation to spermicide use.[72]

While statistical power is a major strength of the case–control approach, power does not assure validity. There are numerous issues that relate to validity (see Chapters 19 and 41); in studies of birth defects, the one that requires particular consideration is recall bias. We previously cited a study of the protective effects of folic acid supplements in

relation to neural tube defects, in which we found different risk estimates among women who reported, at the end of the interview, that they were aware of the hypothesis under study.[39] We believe that this study supports concern about recall bias. As stated above, however, the simple *possibility* of such bias does not necessarily invalidate a case–control study; rather, it requires that investigators consider its existence and make reasonable attempts both to minimize and identify it in their study population.

We cannot review issues in the epidemiologic study of birth defects without alluding to a concern that cuts across all study designs. Birth defects are complex outcomes, and the study of medications in relation to birth defects only adds to this complexity. For all the reasons described in the introduction to this chapter, the rigorous pharmacoepidemiologic evaluation of birth defects requires considerable understanding and experience not only of epidemiology, but also of related disciplines (e.g., pharmacology, embryology, teratology).

The future

There are two important developments that have major implications for the future of pharmacoepidemiologic studies of birth defects. Our knowledge of teratogenicity can be dramatically enhanced by the increasing integration of epidemiology and biology, and it may be diminished by the legal and regulatory climate in which epidemiologists will operate.

Integration of epidemiology and biology

A major frustration among those who conduct epidemiologic studies of congenital malformations is the dearth of understanding of the mechanisms, both structural and molecular, by which defects occur. Advances in molecular biology (such as those related to understanding the role of retinoic acid) will markedly enhance our ability to classify defects in biologically meaningful categories. We may also see advances in the feasibility of using human tissue to identify antenatal drug exposures. Blood and urine have long been available for this purpose, but detection is largely limited to the interval shortly following exposure. For case–control studies in particular, where the mother is typically identified some time after delivery, such sampling is of no value for detecting exposures in early pregnancy. Researchers have explored the usefulness of other tissues, such as meconium[73] and hair[74] in which drugs or their metabolites may persist and accumulate. Although these techniques are still under evaluation and currently lack the ability to estimate timing precisely, they may enable researchers at least to confirm the presence or absence of certain exposures during pregnancy.

With respect to the role of drugs in the etiology of birth defects, there is no question that the most exciting biologic development is the rapid expansion of research focused on genetic factors (see Chapter 34), including those related to variations in fetal drug metabolizing enzymes (DMEs). It has puzzled many that known human teratogens do not produce malformations in all (or even most) exposed fetuses, and many believed that this "incomplete penetrance" was due to differences in host susceptibilities. In 1985, researchers demonstrated that such a phenomenon was likely to account for the inconsistent effect of at least one such teratogen—phenytoin.[75] These workers found that a genetic variant in the detoxification of arene oxide (a radical metabolite of phenytoin) was strongly related to the risk of the major defects associated with phenytoin. Since then, the field has exploded. In anticipation, we and others have added DNA samples to ongoing studies of risk factors for birth defects. Improved understanding of genetic factors will dramatically enhance the identification of subsets of the population who are at increased risk for certain birth defects and the identification of drugs that might warrant particular study.

By analogy with the process of screening for rubella susceptibility or for genetic diseases, one can reasonably look forward to a time when women of childbearing age can be screened for genetic factors which may place them at particular risk for having a malformed infant if they are exposed to a particular drug. Information of this kind has

obvious usefulness in selecting (and avoiding) specific drugs for the treatment of women who are pregnant or at risk for becoming pregnant. (See also Chapter 34 for a more detailed discussion of molecular pharmacoepidemiology.)

The legal and regulatory climate

Whatever their design, studies of exposures in pregnancy in relation to birth defects ultimately depend on the ability to link exposure and outcome information. To accomplish such linkage, researchers require access to information that identifies women who have become pregnant, the outcomes of those pregnancies (including spontaneous and therapeutic abortions), and details as to the presence or absence of malformations. The issue of whether and how such information might be disclosed to researchers has become highly contentious in many countries (see also Chapter 35). For case–control studies in particular, the enrollment of malformed and/or non-malformed subjects requires that hospitals, other health providers, or government agencies make identifying information available to researchers, who then contact eligible subjects in order to invite them to participate in an interview.

At present, there is considerable public anxiety—and even anger—regarding the erosion of confidentiality, especially with respect to financial and related data. Parents of children with birth defects, and particularly those who have undergone therapeutic abortions, are exquisitely sensitive to the disclosure of information on their pregnancies and outcomes. Despite the fact that there is little evidence to suggest that medical researchers have compromised confidentiality, there is a real possibility that epidemiologic research into drug-induced birth defects may be constrained or even eliminated by actions and laws intended to protect confidentiality, without consideration to the substantially different societal roles played by medical research and commercial interests.[76] It is therefore critical that the public be educated about the extent to which epidemiologic research serves the public health, and that they recognize that this benefit can only be accomplished by the provision, under strict controls, of limited confidential medical information to legitimate researchers. At the same time, the public must be reassured—and researchers must accept—that violations of this shared trust will be accompanied by serious penalties. (See also Chapter 35 for a discussion of bioethics in pharmacoepidemiology.)

Steps toward an integrated approach

As noted above, there are a number of broad concerns regarding teratogenesis. The first, both in theory and in practice, is to identify high-risk teratogens (exemplified by thalidomide and isotretinoin); the second is to identify moderate-risk teratogens (exemplified by phenytoin and valproic acid); the third is to provide evidence, where possible, of relative safety. Although not widely appreciated, a combination of the approaches described above can provide much of the information needed to respond to these concerns.[6] Cohorts of exposed subjects, preferably in the form of patient-based pregnancy registries, can identify high-risk teratogens in a timely way. In most instances, the extremely large risks associated with drugs such as thalidomide or isotretinoin tend to overwhelm the distinct methodologic limitations inherent in these approaches. If a drug makes it past that first line of defense, case–control surveillance (or focused case–control studies) can provide the power and rigor necessary to identify whether a drug is associated with specific defects or, if it is not, provide sufficiently narrow confidence limits around the null to provide assurance of relative safety. Sample size considerations will continue to limit our ability to demonstrate safety with respect to relatively rare exposures (or very rare outcomes), but an integrated approach that combines cohort and case–control surveillance offers an effective step towards resolving these two teratogenic concerns.

In the US, such an effort was initiated in 2009. Called the Vaccines and Medications in Pregnancy Surveillance System (VAMPSS),[77] this collaborative approach operates under an infrastructure provided by the American Academy of Allergy, Asthma, and Immunology (AAAAI), and currently utilizes data and analyses provided by two well-established data collection arms—a cohort design that follows women exposed to vaccines or medications of interest, con-

ducted by the Organization of Teratogen Information Specialists (OTIS), and a case–control surveillance arm that includes infants born with the wide range of specific major birth defects (the Slone Epidemiology Center's Birth Defects Study). VAMPSS also includes an Independent Advisory Committee to review the collected data and, when appropriate, to offer recommendations. Although its initial focus is on asthma medications and influenza vaccines and antiviral medications used in the prophylaxis and treatment of influenza in pregnant women, the design of the system provides a highly cost-effective structure that can easily expand to include other vaccines and prescription drugs—as well as OTCs, vitamins/supplements, and herbal products. If sustained, this system can provide, for the first time, a comprehensive and ongoing approach focused on birth defects and the risks and relative safety of all medications taken by pregnant women.

Acknowledgments

I wish to express my appreciation and thanks to Dennis Slone, MD (deceased), Samuel Shapiro, MB, FRCP(E), Carol Louik, ScD, Martha Werler, ScD, Sonia Hernandez-Diaz, MD, ScD, Carla VanBennekom, MPH, and Dawn Jacobs, BSN, MPH, whose advice and close working relationships over the years contributed to the issues covered and views expressed in this chapter.

This work was supported in part by NIH Grant HD46595; over its 34 years, the Slone Birth Defects Study has received grant and contract support from a number of government agencies and pharmaceutical companies.

References

1. Heinonen OP, Slone D, Shapiro S. The women, their offspring, and the malformations. In: Kaufman DW, ed., *Birth Defects and Drugs in Pregnancy*. Littleton, MA: Publishing Sciences Group, 1977, pp. 30–2.
2. CDC. Surveillance summaries: Improved national prevalence estimates for 18 selected major birth defects–United States 1999–2001. *MMWR* 2006; **54**: 1301–05.
3. Gregg NM. Congenital cataract following German measles in the mother. *Trans Ophthalmol Soc Aust* 1941; **3**: 35–46.
4. Lenz W. Thalidomide and congenital abnormalities. *Lancet* 1962; **1**: 45.
5. Warkany J. Problems in applying teratologic observations in animals to man. *Pediatrics* 1974; **53**: 820.
6. Mitchell AA. Systematic identification of drugs that cause birth defects—a new opportunity. *N Engl J Med* 2003; **349**: 2556–9.
7. Stern RS. When a uniquely effective drug is teratogenic: the case of isotreninoin. *N Engl J Med* 1989; **320**: 1007–9.
8. Mitchell AA, Van Bennekom C, Louik C. A pregnancy-prevention program in women of childbearing age receiving isotretinoin. *N Engl J Med* 1995; **333**: 101–6.
9. Mitchell AA, Van Bennekom C. *Isotretinoin Survey*. Presentation at FDA Advisory Committees. Gaithersburg, MD, February 26, 2004.
10. *Risk Evaluation and Mitigation Strategy (REMS).The iPLEDGE Program. Single Shared System for Isotretinoin*. http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM234639.pdf (accessed July 2011).
11. FDA. iPLEDGE. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm094307.htm (accessed Aug 10).
12. Holmes LB. Teratogen update: Bendectin. *Teratology* 1983; **27**: 277–81.
13. Neutel CI, Johansen HL. Measuring drug effectiveness by default: the case of Bendectin. *Can J Pub Health* 1995; **86**: 66–70.
14. Hook EB, Healy KB. Consequences of a nationwide ban on spray adhesives alleged to be human teratogens and mutagens. *Science* 1976; **191**: 566–7.
15. Heinonen OP, Slone D, Shapiro S. Sedatives, tranquilizers, and antidepressant drugs. In: Kaufman DW, ed. *Birth Defects and Drugs in Pregnancy*. Littleton, MA: Publishing Sciences Group, 1977, pp. 335–44.
16. Newman CGH. Teratogen update: clinical aspects of thalidomide embryopathy—a continuing preoccupation. *Teratology* 1985; **32**: 133–44.
17. Lammer EJ, Chen DT, Hoar RM, Agnish ND, Benke PJ, Braun JT *et al*. Retinoic acid embryopathy. *N Engl J Med* 1985; **313**: 837–41.
18. Robert E. [Example of teratogen detection using a birth defects registry: depakine and spina bifida.] *Rev Epidemiol Sante Publique* 1996; **44** (Suppl 1): 78–81 [in French].

19. Whitfield MF. Chondrodysplasia punctata after warfarin in early pregnancy. *Arch Dis Child* 1980; **55**: 139–42.
20. Martin RA, Jones KL, Mendoza A, Barr M, Benirschke K. Effect of ACE inhibition on the fetal kidney: decreased renal blood flow. *Teratology* 1992; **46**: 317–21.
21. 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, National Birth Defects Prevention Study. *Am J Obstet Gynecol* 2011 Apr 21. [Epub ahead of print] PMID: 21514558.
22. Rayburn W, Wible-Kant J, Bledsoe P. Changing trends in drug use during pregnancy. *J Reprod Med* 1982; **27**: 569–75.
23. Werler, MM, Mitchell AA, Hernandez-Diaz S, Honein MA and the National Birth Defects Prevention Study. Use of over-the-counter medications in pregnancy. *Am J Obstet Gynecol* 2005; **193**: 771–7.
24. Broussard CS, Louik C, Honein MA, Mitchell AA and the National Birth Defects Prevention Study. Herbal use before and during pregnancy. *Am J Obstet Gynecol* 2010; **202**: 439.e1–10.
25. Louik C, Gardiner P, Kelley K, Mitchell AA. Use of herbal treatments in pregnancy. *Am J Obstet Gynecol* 2010; **202**: 443.e1–6.
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. Goldsworthy RC, Schwartz NC, Mayhorn CB. Beyond abuse and exposure: Framing the impact of prescription-medication sharing. *Am J Publ Health* 2008: **98**: 1115–21.
28. Daniel KL, Honein MA, Moore CA. Sharing prescription medication among teenage girls: Potential danger to unplanned/undiagnosed pregnancies. *Pediatrics* 2003; **111**: 1167–70.
29. Petersen EE, Rasmussen SA, Daniel KL, Yazdy MM, Honein MA. Prescription medication borrowing and sharing among women of reproductive age. *J Womens Health* 2008; **17**: 1073–80.
30. Fischer MA, Stedman MR, Lii J, Vogeli C, Shrank WH, Brookhart MA, *et al*. Primary medication nonadherence: Analysis of 195,930 electronic prescriptions. *J Gen Intern Med* 2010; **25**: 284–90.
31. Toh S, Mitchell AA, Werler MM, Hernandez-Diaz S. Sensitivity and specificity of computerized algorithms to classify gestational periods in the absence of information of date of conception. *Am J Epidemiol* 2008; **167**: 633–40.
32. Werler MM, Pober BR, Nelson K, Holmes LB. Reporting accuracy among mothers of malformed and nonmalformed infants. *Am J Epidemiol* 1989; **129**: 415–21.
33. Rothman KJ, Fyler DC, Goldblatt A, Kreidberg MB. Exogenous hormones and other exposures of children with congenital heart disease. *Am J Epidemiol* 1979; **109**: 433–9.
34. Bingol N, Fuchs M, Diaz V, Stone RK, Gromisch DS. Teratogenicity of cocaine in humans. *J Pediatr* 1987; **110**: 93–6.
35. Mitchell AA, Rosenberg L, Shapiro S, Slone D. Birth defects related to Bendectin use in pregnancy. 1. Oral clefts and cardiac defects. *JAMA* 1981; **245**: 2311–14.
36. Slutsker L. Risks associated with cocaine use during pregnancy. *Obstet Gynecol* 1992; **79**: 778–89.
37. Bracken MB. Methodologic issues in the epidemiological investigation of drug-induced congenital malformations. In: Bracken MB, ed. *Perinatal Epidemiology*. New York: Oxford University Press, 1984, pp. 435–40.
38. Mills JL, Rhoads GG, Simpson JL, Cunningham GC, Conley MR, Lassman MR, *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.
39. Werler MM, Shapiro S, Mitchell AA. Periconceptional folic acid exposure and risk of occurrent neural tube defects. *JAMA* 1993; **269**: 1257–61.
40. Johnston MC, Sulik KK. Some abnormal patterns of development in the craniofacial region. *Birth Defects Orig Artic Ser* 1979; **15**: 23–42.
41. Hoyme HE, Jones MC, Jones KL. Gastroschisis: abdominal wall disruption secondary to early gestational interruption of the omphalomesenteric artery. *Semin Perinatol* 1983; **7**: 294–8.
42. Werler MM, Sheehan JE, Mitchell AA. Association of vasoconstrictive exposures with risks of gastroschisis and small intestinal atresia. *Epidemiology* 2003; **14**: 349–54.
43. Khoury MJ, Moore CA, James LM, Cordero JF. The interaction between dysmorphology and epidemiology: methodologic issues of grouping and splitting. *Teratology* 1992; **45**: 133–8.
44. Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. *Ann Intern Med* 1979; **90**: 257–61.
45. MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *Lancet* 1991; **338**: 131–7.
46. 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**: 961–8.
47. Martinez-Frias ML, Rodriguez-Pinilla E, Prieto L. Prenatal exposure to salicylates and gastroschisis: a case–control study. *Teratology* 1997; **56**: 241–3.
48. Werler MM, Sheehan JE, Mitchell AA. Maternal medication use and risks of gastroschisis and small intestinal atresia. *Am J Epidemiol* 2002; **155**: 26–31.
49. Khoury MJ, Holtzman NA. On the ability of birth defects monitoring to detect new teratogens. *Am J Epidemiol* 1987; **126**: 136–43.
50. Michaelis J, Gluck E, Michaelis H, Koller S, Degenhardt K-H. Teratogene Effekte von Lenotan. *Dtsch Arztebl* 1980; **23**: 1527–9.
51. Gibson GT, Colley DP, McMichael AJ, Hartshorne JM. Congenital anomalies in relation to the use of doxylamine/ dicyclomine and other antenatal factors. *Med J Aust* 1981; **1**: 410–14.
52. Harlap S, Shiono PH, Ramcharan S, Golbus M, Bachman R, Mann J, *et al*. Chromosomal abnormalities in the Kaiser-Permanente Birth Defects Study, with special reference to contraceptive use around the time of conception. *Teratology* 1985; **31**: 381–7.
53. Czeizel A. Lack of evidence of teratogenicity of benzodiazepine drugs in Hungary. *Reprod Toxicol* 1988; **1**: 183–8.
54. Rebordosa CMG, Kogevinas M, Bech BH, Sørensen HT, Olsen J. Use of acetaminophen during pregnancy and risk of adverse pregnancy outcomes. *Int J Epidemiol* 2009; **38**: 1–9.
55. The National Children's Study. http://www.nationalchildrensstudy.gov/Pages/default.aspx (accessed Aug 2010).
56. Slone D, Siskind V, Heinonen OP, Monson RR, Kaufman DW, Shapiro S. Aspirin and congenital malformations. *Lancet* 1976; **1**: 1373–5.
57. Jick H, Holmes LB, Hunter JR, Madsen S, Stergachis A. First-trimester drug use and congenital disorders. *JAMA* 1981; **246**: 343–6.
58. Jick SS, Terris BZ, Jick H. First trimester topical tretinoin and congenital disorders. *Lancet* 1993; **341**: 1181–2.
59. Ratanajamit C, Skriver MV, Norgaard M, Jepsen P, Schonheyder HC, Sorensen HT. Adverse pregnancy outcome in users of sulfamethizole during pregnancy: a population-based observational study. *J Antimicrob Chemother* 2003; **52**: 837–41.
60. Pasternak B, Hviid A. Use of proton-pump inhibitors in early pregnancy and the risk of birth defects. *N Engl J Med* 2010; **363**: 2114–23.
61. Kuehn BM: Studies probe medication use in pregnancy. *JAMA* 2010; **303**: 601.
62. FDA. *Pregnancy Registries*. http://www.fda.gov/ScienceResearch/SpecialTopics/WomensHealthResearch/ucm251314.htm (accessed July 2011).
63. Dai WS, LaBraico JM, Stern RS. Epidemiology of isotretinoin exposure during pregnancy. *J Am Acad Dermatol* 1992; **26**: 599–606.
64. Chambers C, Braddock SR, Briggs GG, Einarson A, Johnson YR, Miller RK, *et al*. Postmarketing surveillance for human teratogenicity: a model approach. *Teratology* 2001; **64**: 252–61.
65. Chambers CD, Johnson KA, Dick LM, Felix RJ, Jones KL. Birth outcomes in pregnant women taking fluoxetine. *N Engl J Med* 1996; **335**: 1010–5.
66. Chambers CD, Hernandez-Diaz S, Van Marter LJ, Werler MM, Louik C, Jones KL, *et al*. Selective serotonin-reuptake inhibitors and risk of persistent pulmonary hypertension of the newborn. *N Engl J Med* 2006; **354**: 579–87.
67. Yoon PW, Rasmussen SA, Lynberg MC, Moore CA, Anderka M, Carmichael SL, *et al*. The national birth defects prevention study. *Public Health Rep* 2001; **116** (Suppl 1): 32–40.
68. Jick H, Walker AM, Rothman KJ, Hunter JR, Holmes LB, Watkins RN, *et al*. Vaginal spermicides and congenital disorders. *JAMA* 1981; **245**: 1329–32.
69. Mills JL, Harley EE, Reed GF, Berendes HW. Are spermicides teratogenic? *JAMA* 1982; **248**: 2148–51.
70. Shapiro S, Slone D, Heinonen OP, Kaufman DW, Rosenberg L, Mitchell AA, *et al*. Birth defects and vaginal spermicides. *JAMA* 1982; **247**: 2381–4.
71. Louik C, Mitchell AA, Werler MM, Hanson JW, Shapiro S. Maternal exposure to spermicides in relation to certain birth defects. *N Engl J Med* 1987; **317**: 474–8.
72. Warburton D, Neugut RH, Lustenberger A, Nicholas A, Kline J. Lack of association between spermicide use and trisomy. *N Engl J Med* 1987; **317**: 478–82.
73. Lester BM, ElSohly M, Wright LL, Smeriglio VL, Verter J, Bauer CR, *et al*. The Maternal Lifestyle Study: Drug use by meconium toxicology and maternal self-report. *Pediatrics* 2001; **107**: 309–17.
74. Frison G, Vavretto D, Vogliardi S, Terranova C, Ferrara SD. Quantification of citalopram or escitalopram and their demethylated metabolites in neonatal hair samples by liquid chromatography-tandem mass spectrometry. *Ther Drug Monit* 2008; **30**: 467–73.
75. Strickler SM, Dansky LV, Miller MA, Seni MH, Andermann E, Spielberg SP. Genetic predisposition to

phenytoin-induced birth defects. *Lancet* 1985; **2**: 746–9.

76. Andrews EB. International Society for Pharmacoepidemiology. Data privacy, medical record confidentiality, and research in the interest of public health. *Pharmacoepidemiol Drug Saf* 1999; **8**: 247–60.
77. 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: S64–8.

CHAPTER 29
Risk Management

Gerald J. Dal Pan,[1] Stella Blackburn,[2] and Claudia Karwoski[1]

[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, London, UK

Introduction

Risk management is widely used across a variety of settings to identify and assess risks, and to institute measures to mitigate these risks. Such measures can be used to minimize medical errors in health-care settings; limit financial liability in the business sector; minimize or eliminate work-related or recreation-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.

With regard to 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. As in other fields, risk management of medicines is not new, though it has received increased attention in the past two decades, particularly in the past decade. 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, or misused, 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 from 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 United States, 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 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 the sponsor's decisions regarding risk minimization."[1]

In Europe, the concept of risk management is established in legislation. Article 1 (28b) of Directive

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, Fifth Edition. Edited by Brian L. Strom, Stephen E. Kimmel, Sean Hennessy.

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 Europe, risk management incorporates (i) an overview of what is known, and not known about the risks of the drug, (ii) the planning of pharmacovigilance activities to identify additional risks and also characterize the risks, that is seriousness, frequency, risk factors, and public health impact, (iii) measures to minimize the risks including the measurement of the effectiveness, and (iv) a feedback process so that new information about the risks or risk minimization is incorporated into the overview and pharmacovigilance and risk minimization activities adjusted accordingly.

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 indication. Knowledge of a medicine's benefits and risks is developed prior to approval, and refined after approval when its true risk : benefit profile emerges as a result of "real-world" usage. 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 health-care professionals, and the requirement that manufacturers monitor and report to regulatory authorities adverse events that occur with use of the medicine once it is marketed. In the past 20 years or so, additional steps have been taken to manage more actively the risks of certain medications. These measures have included increased communication to patients as well as to health-care 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, health-care organizations and facilities (e.g., hospitals, clinics), manufacturers, and regulatory agencies. Not only does each have a role in ensuring the safe use of a medicine, the interactions among them do as well. 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 health-care system. In this context, the approach to risk management must span the entire lifecycle, be proactive, be scientifically driven, and engage all relevant stakeholders.

Because the risks of medicines can occur at any point in the complex drug use system, managing the risks of medicines requires that the entire drug 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.

The sources of risk from medical products

There are several sources of risks from medical products (Figure 29.1). The known risks of a product are based on prior experience or, in some cases, on the pharmacologic or other properties of

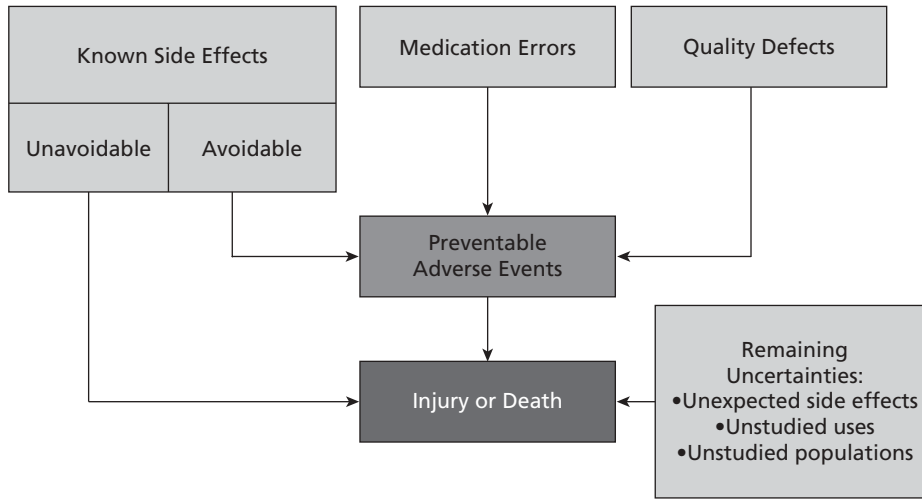


Figure 29.1 Sources of risk from medical products.

the medicine. 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 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 recommendation. Risk management efforts can be used to ensure that preventable adverse events are minimized.

Unavoidable risks are those that occur when all the known necessary conditions for safe use of a product are followed. An unavoidable risk may become a preventable risk once it is identified as a risk but there will always be some risks, which although known cannot be prevented. In these circumstances, risk minimization activities might be directed towards 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 towards regular monitoring of hepatic enzyme levels to identify any hepatic damage as early as possible and so prevent serious hepatitis or hepatic failure.

In addition, risk management efforts can be used to ensure that medications are not administered to patients at higher risk for a serious adverse event, or 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, nor is it achievable. Rather, careful consideration of risk–benefit 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 45) 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]

Because they are, by definition, preventable, medication errors are well suited to risk management efforts. A landmark study published in 2000 estimated that as many as 98 000 people die each year in the United States from medical errors occurring in hospitals.[4] A significant number of those deaths were the result of medication errors. Potential sources of medication errors can include the product's proprietary name (if it is similar to the name of another medicine, especially if the two medicines have other similar characteristics), the established name, and the design of the container, carton, and packaging. Errors can also occur if the product labeling for health-care professionals or patients is not clear.

Because medication errors can occur anywhere in the medication use chain, efforts to minimize the risk of medication error must involve multiple stakeholders. In the United States, FDA reviews proposed proprietary names of medications to ensure that these names are not similar in spelling or pronunciation to the proprietary or established names of other medicines.[5,6] In addition, FDA reviews the proposed carton, container, and packaging to ensure that these do not have features that could lead to medication errors. Many professional organizations also have recommendations to minimize medication errors.

Similarly in Europe, medicines authorized through the centralized procedure have their invented (or brand) name approved by the (invented) name review group to check that there are no products licensed with similar names in Europe, which could lead to confusion. In the centralized procedure, the European Commission (based on a scientific assessment by the Committee for Medicinal Products for Human Use, CHMP) issues one license which covers 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.

Product quality problems occasionally result in adverse events.[7,8] These problems are unusual in both the United States and Europe because of the great attention paid to product quality control and quality assurance during manufacturing. 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 lifecycle of a medicine, as discussed below.

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 43). 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 integrating 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 lifecycle 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 lifecycle

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., geriatric patients), side effects may be discovered if these groups are treated with a product after it goes on the market. 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 1992 and 2006 indicated that safety-related labeling changes were being made as long as 12 years after the products were approved.[9] Because of this lifecycle 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 medications.

Risk management applies to all medicines

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, many antineoplastic agents have agranulocytosis as a common side effect. This life-threatening side effect must be carefully managed in order for the potential benefits of the drug to outweigh this risk. At the other end of the spectrum, many topical over-the-counter 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 stopped. 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 but serious side effects are acute liver failure, aplastic anemia, torsade de points, 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.

Over-the-counter medicines are drugs that have been found to be safe and appropriate for use without the supervision of a health-care professional such as a physician, and they can be purchased by consumers without a prescription. Most over-the-counter medicines are for symptomatic relief of conditions that consumers can diagnose and manage themselves. When these medications are taken properly, many of their side effects are generally mild. However, there can be serious, even life-threatening or fatal, side effects of over-the-counter medications when they are not taken properly. For example, acetaminophen (paracetamol), one of the most widely used over-the-counter analgesics, is a generally very safe over-the-counter medication 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 United States.[10]

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 non-steroidal anti-inflammatory agent, were 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. 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.[11] 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 allows manufacturers to work with regulators on risk management planning during the drug development phase. A proactive approach in the postapproval phase demands that manufacturers, regulators, and practitioners have in place systems 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 also be used in proactive ways to learn about the safety of a medicine in as efficient a manner as possible. A carefully designed risk management plan can identify risks, manage risks, communicate risks, and assess the effectiveness of these efforts in a proactive way. Like the lifecycle approach noted above, the proactive nature of risk management planning demands the constant vigilance of all stakeholders.

Risk management is an iterative process involving multiple related 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 inter-related activities. In broad categories, these activities include risk assessment, risk mitigation, evaluation of risk mitigation measures, and risk communication. These activities occur throughout the product's lifecycle, 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.

Risk assessment

Risk assessment consists of identifying, characterizing, and quantifying the risks associated with the use of a medicine. 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. 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 genotoxicity, carcinogenicity, immunotoxicity, and reproductive toxicity. 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.

Clinical pharmacologic studies establish the pharmacokinetic profile of the medicine, 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 medication, 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 medication. Depending on the number of subjects exposed prior to approval, less common adverse events can also be detected. It is important to pay careful attention to the design of the preapproval safety program in order 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 medications, 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.

Because even large clinical development programs cannot identify all risks associated with a product, it is imperative that risk assessment continues in the postapproval period, when large numbers of persons will be exposed to the medicine, including many with co-morbid conditions or on concomitant medications not present in clinical trials. Postapproval risk assessment can be based on either non-experimental data or on clinical trial data. Non-experimental 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 medication usage. The use of many of these data sources, and the methods underlying their use, are covered in other chapters of this book (see Chapters 10–18), and will not be covered 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 professionals, introduction of checklists or monitoring requirements to prevent inappropriate prescription, restrictions on the use of the medicine, or, rarely, market withdrawal.

Risk assessment of medicines, both in 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 (see also Chapter 45), and the potential for medication errors, throughout the product's lifecycle. The identification and assessment of medication errors is different in some ways from the identification of adverse drug reactions. The identification of a medication error generally requires that someone report that an error has occurred, though the initial report may often 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. Because 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. The identified reasons and causes may relate to certain characteristics of the product itself, to the larger medication use system in which the product is used, or to an interaction of the two. 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.

Risk mitigation

Risk mitigation refers to a set of activities designed to minimize the risks of a medicine while preserving its benefits. 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 drug 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 medications are too dangerous to be used without the involvement of an appropriately qualified health-care professional, whose judgment can be used to ensure that, for a particular patient, the potential benefits outweigh the potential risks.

For most prescription medicines, the prescription status, professional label, and information directed to patients are sufficient risk minimization measures to ensure that the benefits of a medicine outweigh its risks. In some cases, additional measures are needed. These additional measures are usually designed to address one, or at most a few, specific serious risks. While specific measures may differ from one country or region to the next, these measures generally fall into one of three categories: (i) focused information for patients; (ii) additional, focused information for practitioners; and (iii) measures that restrict, in some way, the use of the medicine. A risk mitigation strategy may use one or more of these measures.

Because additional risk mitigation measures are focused on only one or a few serious risks, the communications focused both toward health-care professionals and patients must address these risks in detail, while at the same time putting these specific serious risks in the context of the overall risk-and-benefit profile of the medicine. The details of the communication depend on the nature of the risk and the specific steps that can be taken to mitigate it. 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 if the potential benefits outweigh the potential risks in their individual situation. In other cases, the communication focuses both on the nature of the risk as well as on specific steps that can be taken to prevent it or to recognize it early.

The extent to which specific information about the risk needs to be communicated will depend upon the context in which it is used. For example, specific risk management for a drug that 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 it might be appropriate for additional information about the need for periodic monitoring of electrocardiograms, risk of concomitant use with other QT interval prolonging drugs, etc., to be provided as a risk minimization activity.

The design of measures to restrict, in some way, the manner in which a medicine is prescribed or used is complex and, more than other risk mitigation efforts, requires complex interactions with the medication distribution and use system. The details of how this is done in the United States are described later in this chapter.

Because these additional risk mitigation measures are highly focused, it is critical that they have clearly specified and well thought out goals. In the absence of clear goals, proper interventions cannot be designed, and the impact of these interventions cannot be measured.

Evaluation of risk mitigation measures

Evaluation of risk mitigation activities is a critical component of a risk management system, and is also a relatively new endeavor in the context of the medication use system. The evaluation of a risk mitigation activity is closely related to the risk assessment activities, but it also differs in some ways. While traditional risk assessments are designed to identify, characterize, and quantify previously unknown risks, evaluations of risk mitigation activities are designed to examine the impact of these risk mitigation activities. It is thus important that the measures used to evaluate a risk mitigation activity focus on the goals of the risk mitigation plan.

Evaluation of risk mitigation measures can occur at several levels. There are many frameworks that can be used to describe these levels. One such framework is to evaluate risk mitigation activities on three levels—process, behavior, and outcome. 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 taken to minimize a particular risk, the process component of the assessment could consist of determining the proportion of patients who receive the information. Second, the evaluation can determine if certain behaviors are being followed. In the above example, measurements of behavior could assess if patients read the information they are given, if they understand the information, and if they do the specific things the information recommends. Third, the evaluation can measure the frequency of the health outcome that the risk mitigation strategy is designed to minimize. These three types of evaluation are quite different from each other, in terms of both the data needed to conduct the evaluations as well as the methods to analyze the data. In some cases, evaluations of adherence to specific processes may be easier to conduct than other types of evaluations. However, it is important that the final health outcome of interest be evaluated, since adherence to a process may not guarantee attainment of the health outcome of interest.

Assessing the risks of a medicine, instituting risk mitigation measures, and evaluating the impact of those measures is an iterative process. As new risks are identified, new risk mitigation measures may need to be put into effect. These new measures will then need to be evaluated, and the risk mitigation measures may need to be modified. This iterative process occurs throughout the lifecycle of the product. Pharmaceutical manufacturers are held responsible for the safety of their products, so it is usually they who fund risk assessments and put in place risk mitigation procedure evaluations. Regulators review the results of manufacturers' testing, proposed risk mitigation strategies, and the evaluations of the 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 independent of manufacturers and regulators or in collaboration with them.

Defining and setting the goals of a risk mitigation strategy presents several challenges, especially

because it is critical to define the overall success or failure or a risk mitigation strategy in terms of the actual health outcomes of interest. First, data on the actual frequency of the health outcome of interest may not be readily available or may not be reliable or representative of the entire population taking the medicine. Second, without prior experience with similar risk mitigation strategies, it is difficult to set a quantitative goal for the risk mitigation strategy. While the aspirational goal for mitigation of certain avoidable risk is zero occurrences of the event, the complexities of the medication use system, and the reliance on the behavior of so many individuals, may make such a goal impractical.

For example, if a drug is a known human teratogen, the aspirational goal may be no fetus exposed to the drug. However, no pregnancy prevention measure has a 100% success rate and humans are fallible, so a more practical goal may be no children born with birth defects. In this case, the strategy would include the provision of information about the teratogenic effects of the drug, identification of patients particularly at risk of pregnancy, a pregnancy prevention plan, measures to identify any pregnancy as soon as possible, and rapid access to experts to assess the effect of any fetal exposure.

Setting a true numeric goal is difficult. While prior experience may yield a reasonable estimate of the prestrategy frequency of occurrence, setting a realistic poststrategy goal is generally not an intuitive process. In the above example, termination of pregnancy may not be acceptable to some, so a goal of no children born with birth defects may also not be achievable. But equally, setting a goal of fewer than five children born with birth defects may be ethically difficult to justify.

Currently, the performance of many risk management systems is not well studied and not well understood. This is true for the system as a whole, as well as for the components of the systems. Interventions that are implemented because they are thought to be effective may turn out not to be effective. The effectiveness of risk management interventions is not well understood, and the characteristics of effective risk management systems are not well catalogued. These limitations in knowledge, coupled with the fact that many current systems are individually designed, make it difficult to implement new systems with confidence that they will be effective.

Risk management systems need to be assessed, but current methods of assessing risk management systems are not well developed. Specifically, methods to measure the impact of risk management systems and their component parts on processes, behaviors, and most importantly health outcomes, need to be developed, so that effective measures can be continued and ineffective ones can be removed.

Some aspects of the risk management system, especially those that impose strict restrictions on the use of a medicine, may be burdensome on the health-care 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. Obtaining such data will be challenging, because it involves 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.

Risk communication

Communicating information about the benefits and risk of medicines is central to managing the risks of these products. Risk communication is a broad field, and a full discussion is beyond the scope of this chapter (see Chapter 43). Communication has traditionally been directed towards health-care professionals, but in recent years increasing attention has been paid to communications directed towards patients and consumers.

The principal form of communication to health-care professionals is the product's approved professional labeling, which is designed to present to the health-care professional information needed to prescribe the product in a way that the potential

benefits outweigh the potential risks. In Europe, 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 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 stopping the medicine or changing the dose. Third, the label contains information about the conditions of safe use of the medication, such as the proper dosing (including, when applicable, the dose adjustments needed for renal and hepatic impairment or those based on age), drug–drug interactions, drug–disease interactions, use in pregnant or lactating women, and use in other specific clinical situations. Labeling directed to 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 non-prescription medicines. However, as noted in the example of cisapride in Chapter 8, changes in labeling do not always result in changes in prescribing practices.

Additional communications to health-care professionals come in the form of so-called "Dear Healthcare Professional Letters" or "Dear Doctor Letters". 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 is often included. The letter can also highlight actions that the health-care professional can take in prescribing and dispensing the product, as well as other measures that can help assure the product's safe and appropriate use. Full prescribing information is generally attached, so that prescribers can put the new information into context. Frequently, in Europe, the text of these letters is agreed with the appropriate regulatory authority and, in some cases, its provision to health-care professionals made a condition of the marketing authorization.

Information to patients can come in a variety of forms. One common form is product-specific information directed towards patients. This can take the form of approved patient labeling, which is developed by the manufacturer and reviewed and approved by FDA. Approved patient labeling includes the Medication Guide or a Patient Package Insert. Medication Guides are used when there is a need to communicate certain safety information, or when certain conditions of safe use must be highlighted.[12] By regulation, FDA may require that Medication Guides be issued with certain prescribed drugs and biological products when the Agency determines that one or more of the following circumstances exists:

- 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 is specified in regulation, and distribution requirements are set forth in regulation.[12]

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.

In Europe, all medicines are required to have a package leaflet (sometimes also referred to as a patient information leaflet) which must be provided to the patient. This leaflet is based upon the information provided to the physician (the SmPC) but written in patient-friendly language. There is a requirement to test the readability of the package

leaflet with an appropriate target group of patients/consumers and provide the results to the competent 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 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 State.

In addition to the mechanisms listed above, manufacturers use a variety of other communication tools to reach practitioners and patients. These can include print and broadcast advertising, websites, and other communications that use 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 mitigation strategy. In Europe, advertising of prescription-only medicines directly to patients is prohibited.

Consumer Medication Information is an alternative form of drug-specific patient information used in the United States. Unlike approved patient labeling, Consumer Medication Information is neither developed by the product's manufacturer nor is it regulated by 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.

A recent study of Consumer Medication Information highlights the challenges in developing effective information for patients and consumers.[13] This study, which used professional shoppers (i.e., persons paid by the researchers to simulate patients seeking to have prescription filled) to fill prescriptions for lisinopril and metformin in a national sample of 365 pharmacies, evaluated the information according to explicit criteria. The study found that 94% of pharmacies provided written information for these products, but that there were notable shortcomings regarding comprehensibility, legibility, and directions for use. There was a wide range of length among the samples of written information, even amongst the samples that scored well on content criteria. The authors concluded that further research on the best methods of presenting information about medicines to patients is needed.

Regulatory agencies have also been engaging in increasing efforts to communicate the risks of medicines. For example, in March 2007, FDA issued a guidance document entitled *Drug Safety Information—FDA's Communication to the Public*.[14] In it, FDA notes its intent to disseminate information on "emerging drug safety information", which it defines as "information FDA is monitoring or analyzing that may have the potential to alter the benefit–risk analysis for a drug in such a way as to affect decisions about prescribing or taking the drug (i.e., an important drug safety issue), but that has not yet been fully analyzed or confirmed." In establishing this system, 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 hand. Because analyses and interpretations of drug safety information are often not clear cut, communication of findings requires a balanced and impartial approach.

In Europe, the national regulatory authorities communicate drug safety information using different methods, which depend upon what has been established for the individual country. The European Medicines Agency (EMA) plays a central role in coordinating the communications. When a drug safety issue has been discussed and agreed by the Committee for Human Medicinal Products (CHMP), the 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 CHMP for a scientific Opinion, then the start and reasons for the referral are also announced as well as the final conclusions.

The EMA also coordinates the release of information in the 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). The EPAR is a scientific document which summarizes the information that the CHMP evaluates in giving their 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 lifetime of the medicine. In addition, the SmPC, the package leaflet, the labeling, and the conditions for the authorization also appear on the website and similarly are updated throughout the lifetime of the medicine.

The requirement for an assessment report to be publically available also applies to medicines authorized by other routes (national, mutual recognition, and decentralized procedures), but the exact format in which it appears will vary.

Methodologic problems to be addressed by pharmacoepidemiologic research

The 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 medication. For most prescription medications, 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 very 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 medication or after a long latency period. In addition, real-world patient populations include patients with a broader range of comorbidities, on a wider variety of medications, and more severe underlying disease than those included in clinical trials. In practice, patients may be treated with dosing regimens or may receive the medication 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 United States in August 2009 for the treatment of infantile spasms and for the treatment of refractory complex partial seizures in adults. It had been already approved in the United Kingdom. Several years after its approval in the UK, 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 its approval in the US, the sponsor conducted, at the request of European Medicines Agency, a study to characterize better the occurrence of peripheral visual field defects. The primary endpoint was formal visual field testing every 4 to 6 months for 3 years. This measure allows for 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. Amongst adults 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.[15] These data demonstrate that clinical trials can be used to characterize and quantify specialized drug safety questions. In the United States, the product was approved in August 2009 with a Risk Evaluation

and Mitigation Strategy (REMS) to address this serious safety concern, as described later in this chapter.

An important use of pharmacoepidemiology is to measure how medications are used in practice, especially if they are used under conditions that can lead to adverse outcomes (see also Chapter 24). Examples of pharmacoepidemiologic findings that could signal that a product is not being used appropriately include a finding that a medication is being prescribed concomitantly with a contraindicated medication, 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 medication 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 health-care data, especially those with longitudinal patient-level data, are often useful.

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, amongst 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 long-acting beta agonists are used in conjunction with inhaled corticosteroids 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.[16] 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).[17] 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. FDA has recently required new labels for LABA products and will assess changes in prescribing patterns based on data documenting drug use.[18]

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 heath databases, as described below, 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. Generally, these databases are more useful for characterizing and quantifying known drug risk, rather than identifying new risks. 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 medications in the United States, available in several single-ingredient and multi-ingredient, over-the-counter, and prescription products. Although acetaminophen is generally safe when used as directed, misuse and overdose can cause acute liver failure, sometimes resulting in transplantation or death. Overdoses can be either intentional or unintentional. To provide context for risk mitigation activities, Nourjah and colleagues at FDA examined several national databases to quantify this problem.[19] 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 both the overall magnitude of acetaminophen overdoses in the United States 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. In the US, the FDA is considering what steps may be taken to minimize the occurrence of serious liver damage related to acetaminophen use.

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 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 noted above, 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 the third level (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 (i.e., the specific risk is not mitigated). 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. In the final analysis, a risk mitigation strategy is successful only if there is mitigation of the specific risks that are the focus of the strategy.

The example of an analysis of risk mitigation strategy for isotretinoin illustrates how pharmacoepidemiology can be used to assess the impact of various program measures on health outcomes. Isotretinoin is a medicine that is uniquely effective in the treatment of severe, recalcitrant nodular, cystic acne. Approved in the United States in 1982, it also can cause severe birth defect and intrauterine fetal deaths. To minimize exposure during pregnancy, the manufacturer, in 1988, developed the Accutane Pregnancy Prevention Program. By 2000, FDA had concluded that the Pregnancy Prevention Program was not effective in minimizing exposure during pregnancy. The manufacturer then developed the System to Manage Accutane-Related Teratogencity (SMART). An essential feature of this program was a qualification sticker that was to be attached to a prescription for isotretinoin, which was to indicate adherence to certain program-mandated steps by prescribers and female patients. Prescribers were required 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 continuing medical 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 effect 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 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 7 days before the actual start of treatment. When each of these steps had been met, the patient was "qualified", and was to present a 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 negative serum and urine pregnancy tests.

Prior to implementing the SMART program, the manufacturer and FDA agreed to evaluate the program's effectiveness during the first year.[20] A Pharmacy Compliance Survey found that, after the third month of the program, compliance with the requirement for a sticker was generally high, above 99% for urban pharmacies and above 90% for rural pharmacies. The proportion of stickers that contained information on patient gender and qualification date was similarly high. A 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 age 15 through 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 the product. The above analysis shows that despite reasonably high compliance with the placement of a 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, performance of pretreatment pregnancy tests and use of birth control. These findings indicate that process measures that are a surrogate for clinical events need to be validated.

Another role for pharmacoepidemiology is in the area of assessing risk communication. The assessment of communications 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 patients-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 measureable health outcomes.[21] To assess the relationship of various communication strategies to health outcomes, Shrank and colleagues took advantage of a deliberate effort of one large pharmacy chain to improve its patients labeling.[22] They then used administrative claims data from one insurance carrier in two states in the United States to look at various health outcomes for patients to whom outpatient medications for one of nine chronic conditions were dispensed. Because these data contained detailed information on the specific pharmacy at which the medications 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 that 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% CI: 0.15–1.86) or inpatient and emergency department care (event rate ratio: 0.88, 95% CI: 0.62–1.24) amongst 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 United States

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.

Section 901 of the Food and Drug Administration Amendments Act of 2007 amended the Food, Drug, and Cosmetic Act by granting FDA, amongst other authorities, the authority to require manufacturers, under certain circumstances, to develop and implement a Risk Evaluation and Mitigation Strategy (REMS). While REMS are an important part of FDA's risk management strategy, they are not the only mechanism the Agency uses. REMS in the United States are additional risk mitigation measures that go beyond the prescription status of the medicine, the requirement for professional labeling, and the occasional use of patient labeling.

A REMS is a required risk management plan that utilizes tools beyond routine labeling to ensure that benefits of a drug outweigh its risks. Title IX, Subtitle A, Section 901 of the Food and Drug Administration Amendments Act (FDAAA (PL 110-85)) amended the Federal Food, Drug, and Cosmetic Act to authorize the FDA to require application holders to develop and comply with REMS if specific statutory criteria are met. These provisions became effective on March 25, 2008. The new 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 initial approval of an application, FDA offices responsible for review of the drug or biologic and for postapproval safety review determine whether a REMS is needed to ensure that benefits of the drug outweigh its risks. FDAAA (PL 110-85) requires FDA to consider certain factors (listed in Table 29.1) in making the determination. In the postapproval phase, FDA may determine that a REMS is needed if FDA 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

Table 29.1 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).

literature, or other scientific data deemed appropriate by 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.

All REMS for NDAs and BLAs must include a timetable for submission of assessments of the REMS and FDA must determine which of the following additional elements will be included in a REMS, if criteria specified in the law are met:

- a Medication Guide (MG) or a Patient Package Insert (PPI)
- a communication plan
- Elements to Assure Safe Use (ETASU)
- an implementation system.

A Medication Guide may be required as part of a REMS to inform patients about serious risks associated with the product and may also be used to provide patients with information necessary for the safe use of the product. REMS with communication plans are designed to ensure that health-care providers are made aware of important information for safe use of the drug. Elements to assure safe use (listed in Table 29.2) 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. 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 health-care providers and other participants.

The minimal timetable for submission of assessments of a REMS includes assessments by 18 months, by 3 years, and in the seventh year post-REMS approval. FDA may require more frequent assessments that may be specified in the REMS. The assessments can be eliminated after 3 years if FDA determines that serious drug risks have been adequately identified, assessed, and are being adequately managed.

FDAAA (PL 110-85) specified that drugs approved before September 27, 2007 with ETASU were deemed to have REMS and were required to submit proposed REMS for approval by September 21, 2008. FDA identified 28 product applications deemed to have in effect an approved REMS.[23]

Table 29.2 Risk evaluation and mitigation strategies in the US: elements to assure safe use

A. Health-care providers who prescribe the drug have particular training or experience or are specially certified
B. Pharmacies, practitioners, or health-care settings that dispense the drug are specially certified
C. The drug be dispensed to patients only in certain health-care 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).

Regulatory framework in the European Union

The requirement for risk management in the European Union (EU) is specifically included in legislation. In Europe, the term "medicinal product" is also defined in the legislation and the term includes both chemical and biological 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. Article 8 (3)(ia) of Directive 2001/83/EC requires the inclusion of: "a detailed description of the pharmacovigilance, and where appropriate, of the risk management system which the applicant will introduce."

The terms "*pharmacovigilance systems*" and "*risk management systems*" mentioned in the legislation frequently cause confusion. A pharmacovigilance system refers to the measures that a company puts in place to meet the requirements in the legislation for pharmacovigilance. These requirements include the need to have within the company an "appropriately qualified person responsible for pharmacovigilance." This person must reside within "the Community" and so is known as the EU Qualified Person. The pharmacovigilance activities that (s)he

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 and 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 guidances;
- 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 qualified person, and the various processes, standard operating procedures (SOPs), etc. by which an individual company ensures compliance with pharmacovigilance legislation. The requirements for the description of the pharmacovigilance system are described in a chapter of Volume 9A of the Rules Governing Medicinal Products in the European Union.

Whereas the pharmacovigilance system applies to the people and processes in a whole company, a risk management system is usually product specific. It describes the specific risks pertaining to a particular product, how they will be identified, investigated and characterized further, and the risk minimization activities which will be put in place to prevent or mitigate them.

The situations which constitute "*where appropriate*" and the specific requirements for the information to be included in the "*detailed description….of the risk management system* " are laid down in the Guideline: Requirements for risk Management Systems for Medicinal Products for Human Use, which forms Chapter I.3 of Volume 9A of The Rules Governing Medicinal Products in the European Union. The guideline states that the requirement for a description of the risk management system can be fulfilled by the submission of an EU risk management plan. The amendment of Directive 2001/83/EC, which introduced risk management, came into effect in November 2005 so formal risk management plans for some products have existed in Europe since this date.

The guideline, which specified how the legislation relating to risk management plans would be put into practice, also implemented the International Conference on Harmonisation (ICH) E2E Guideline on Pharmacovigilance Planning. The ICH guideline describes how the safety profile of a medicine should be discussed and summarized within a Safety Specification. This Safety Specification is a synopsis of the entire development program, and any postauthorization experience with the medicine, and ends with the identification of safety concerns. These safety concerns relate to important identified risks, important potential risks, and areas where important information is missing. How these safety concerns will be investigated further is discussed in the Pharmacovigilance Plan. The information in the Safety Specification, and hence the Pharmacovigilance Plan, is updated throughout the product's life cycle.

The Safety Specification and Pharmacovigilance Plan from ICH E2E, with a few additional EU specific items in the Safety Specification, form the first part of the EU-RMP, as shown in Table 29.3. Part II of the EU-RMP consists of an "Evaluation of the need for risk minimization activities" and, if needed, a formal risk minimization plan. Therefore in Europe, risk management plans effectively consist of complimentary parts:

- what is known and not known about the safety profile of a medicine;
- how the safety concerns will be investigated, other risks identified, and all characterized;
- what risk minimization measures are needed and whether a specific risk minimization plan is necessary.

Table 29.3 The EU risk management plan

Part I	
Safety specification Pharmacovigilance plan	} ICH E2C
Part II	
• Evaluation of the need for risk minimization activities	
If a need for additional risk minimization activities:	
• Risk minimization plan	

Routine risk minimization

Certain risk minimization activities in Europe are regarded as "routine" as the legislation in Directive 2001/83/EC requires them for all medicines. A risk minimization plan is only needed if there is a need identified for what are termed "additional" risk minimization activities, that is activities in addition to the routine risk minimization required for all medicines.

These "routine" risk minimization activities are:

- legal status of a medicine (i.e., whether subject to medical prescription or not subject to medical prescription);
- authorized pack sizes;
- summary of product characteristics (information for health-care practitioners);
- package leaflet (information for patients);
- labeling (the immediate and outer packaging of the medicine).

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.

Medicines within the "restricted" medicine category will have the details of the restriction in the SmPC. One of the complications within Europe is that the definition of a "specialist" is not the same across Europe and likewise the health-care setting is not common across Europe. For this reason, the phrase "physicians with expertise in the management of X" is often used to define the area of medicine the physician should be working in and details of the equipment necessary, for example "in a setting where resuscitation equipment is immediately available" will be used, instead of defining use in a hospital, clinic, or surgery.

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. They are thus legally binding on the Marketing Authorization Holder (MAH).

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 they give 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 usually be adopted by the European Commission and form part of the legally binding conditions of the Marketing Authorization. In these circumstances, the European Commission will also adopt a Decision (under Article 127 (a) of Directive 2001/83/EC) which is directed to the Member States and requires them to ensure that the conditions and restrictions are applied in their country.

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 CHMP (and the European Commission) to be a legally binding condition of the marketing authorization with which the MAH must comply. Because there are differences in national legislation, in theory the conditions or restrictions of the Marketing Authorization can contain provisions that cannot be implemented legally in all Member States although the medicine is actually authorized for use across Europe. In these circumstances the MAH would not be able to sell the medicine in the particular country. In practice, where possible, this situation is prevented by careful wording of the conditions and restrictions.

As a result, the conditions and restrictions are frequently the subject of much debate at CHMP since not only is the practice of medicine different across the EU, the customs and mores of the coun-

tries are very different. Therefore, something which is legally possible may not be socially acceptable in a particular country or may be contrary to normal medical practice. Usually, the required risk minimization can be achieved by carefully dividing the requirement into smaller steps or processes which can be implemented in various ways to suit the particular situation in the different countries. Therefore, in the most difficult situations, the marketing authorization may have multiple conditions which make up the additional risk minimization activities—all of which must be complied with by the MAH.

Each condition will usually state what must be achieved but how this will be done will remain flexible. For example, the conditions of the Marketing Authorization may specify that the MAH shall set up a controlled distribution. Because of the differences in the way that health care is delivered in each of the Member States, there are at least four different ways in practice in which distribution is controlled across Europe. 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 it 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 is 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 MAH, 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 of 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 had centers involved in clinical trials, training centers did not exist initially in all countries. As a consequence, training of the 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 whilst still allowing flexibility in how this is achieved.

The ability to set conditions and restrictions does not only apply to medicines at the time of authorization. If during the lifetime 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, it is possible to remove the conditions or restrictions.

Risk management example in the US—vigabatrin

Vigabatrin is an antiepileptic medication that can cause bilateral peripheral visual field constriction in a high percentage of patients. An oral solution formulation is indicated as monotherapy for pediatric patients with infantile spasms for whom the potential benefits outweigh the potential risk of visual loss. An oral tablet formulation is indicated as adjunctive therapy for adult patients with refractory complex partial seizures who have responded inadequately to several alternative treatments. Because of the potential for visual loss, described earlier in this chapter, vigabatrin was approved in

the US with a REMS. The goals of the REMS, which are directed at managing the risks of visual loss, are:
1 to reduce the risk of a vigabatrin-induced vision loss while delivering benefit to the appropriate patient populations;
2 to ensure that all patients receive a baseline ophthalmologic evaluation; 50% of patients will receive this within 2 weeks of starting vigabatrin and 100% within 4 weeks;
3 to discontinue vigabatrin therapy in patients who experience an inadequate clinical response;
4 to detect vigabatrin-induced vision loss as early as possible;
5 to ensure regular vision monitoring to facilitate ongoing benefit–risk assessments;
6 to inform patients/parent or legal guardian of the serious risks associated with vigabatrin, including vision loss and increased risk of suicidal thoughts and behavior.

The approved REMS consists of five main components. The first is communication to patients through a Medication Guide. The REMS also specifies that the physician is to review the Medication Guide with the patient, parent, or legal guardian prior to starting vigabatrin therapy. The Medication Guide focuses on the potential for visual loss and visual damage, including a description of some of the symptoms that may indicate a visual field defect and instructions to seek medical attention should these symptoms arise. The Medication Guide also includes other important information about the safe use of the product. A second component of the REMS is a communication plan directed toward health-care providers. In this case, the communication takes the form of Dear Healthcare Professional Letter, and is directed at ophthalmologists. The letter describes the indications for vigabatrin, the patterns of visual loss seen in some patients treated with vigabatrin, the mandatory vision monitoring for adults and infants treated with vigabatrin, the challenges of assessing visual field defects in infants, brain magnetic resonance imaging findings in patients treated with vigabatrin, and an overall description of the REMS program. The third component of the REMS encompasses "elements to assure safe use", which function as a type of restricted distribution system. The particular elements in this case include special certification of health-care providers who prescribe vigabatrin, special certification of pharmacies that dispense vigabatrin, a requirement that vigabatrin be dispensed to patients with evidence or other documentation of safe-use conditions, and enrollment of each patient using vigabatrin in a registry. The conditions of safe use stated in the REMS also require that the patient, parent, or guardian indicate that: (i) they have read the Medication Guide; (ii) the prescriber has explained the risk of visual loss; (iii) they understand that the visual loss is irreversible; (iv) they understand that prescribed vision assessments are required; (v) they understand that periodic vision assessments are required for the duration of treatment and after treatment is discontinued, although it does not protect against visual loss; and (vi) they understand that the response to vigabatrin will be assessed after a short trial period and that patients with insufficient responses will discontinue using vigabatrin. Specialized forms to document these components of this element are part of the REMS. A fourth component of the REMS is an implementation system, which describes how operational elements and responsibilities specified in the REMS will be implemented by the drug company that markets vigabatrin. A fifth element is a timetable for submission of assessments. For vigabatrin, assessments were required every 6 months from the date of approval of the REMS for 1 year, and annually thereafter.

The vigabatrin REMS Assessment Plan comprises a number of components including a plan to assess patients'/caregivers' and health-care providers' understanding about the risks and safe use conditions of vigabatrin, an assessment of whether vision monitoring is being adhered to, and an assessment of the rate of visual field events.

The vigabatrin REMS underscores certain important features of modern pharmaceutical risk management. First, the goal of a risk mitigation strategy is not always a zero rate of serious adverse events. While the complete prevention of a serious adverse event is obviously desirable, such a goal is not always obtainable. Yet, here are patients for whom the risk of a medication outweighs its benefits. In these cases, it is important for health-care

providers to understand well the nature of the risk and to ensure that the proper pre- and post-treatment monitoring is done. It is equally important they be able to identify those patients for whom the potential benefit outweighs the potential risk and to explain these risks to the patients. Second, the target of communicating the risks of a medication is not only those physicians who will prescribe the medication. It is equally important to communicate these risks to specialists who may see the complications of treatment. In the case of vigabatrin, neurologists would be the specialty group to target if one wants to direct messages to prescribers. However, ophthalmologists need to be aware of the risk of bilateral concentric peripheral visual field constriction, since they comprise the specialty group that evaluates this adverse event. Third, a robust plan must be put into place at the time the risk mitigation strategy is implemented, to determine whether the strategy is effective in mitigating the risk.

Risk management example—eculizumab

Eculizumab is a complement inhibitor for the treatment of patients with paroxysmal nocturnal hemoglobinura (PNH) to reduce hemolysis. It is given by intravenous infusion once a week for 4 weeks, then every 2 weeks thereafter. Eculizumab can cause meningococcal infection because it inhibits terminal complement, making patients more susceptible to infection with *Neisseria meningitides*. Persons who have deficiencies in the terminal common complement pathway (C3, C5–9) are among the groups of individuals that are at increased risk for acquiring meningococcal disease.

Risk management approach for eculizumab in the US

Because of the potential for meningococcal infection, eculizumab was approved in the US with a boxed warning regarding the increased risk of meningococcal infections, the need to vaccinate patients with a meningococcal vaccine, to monitor patients for early signs and of meningococcal infections, evaluate, and treat with antibiotics if necessary. At the time of US approval, the product manufacturer committed to developing and implementing a Risk Minimization Action Plan (RiskMAP) which included an educational plan intended for health-care providers.

Following approval, the FDA became aware of postmarketing information that indicated a delay in diagnosing a case of meningococcal meningitis. The health-care professional that prescribed eculizumab was unaware of the need to educate the patient regarding the risk of meningococcal infection. In late 2008, the FDA issued a letter to the company notifying them to submit a REMS for eculizumab. The REMS was approved in June 2010. The goals of the REMS, which are directed at managing the risks of meningococcal infection, are:

- to limit the occurrence and morbidity associated with meningococcal infections;
- to mitigate serious outcomes for patients who develop infection with *N. meningitidis* and other systemic infections;
- to impart important safety information before initiating treatment with eculizumab and ensure proper use of eculizumab while patients remain on therapy by:
  - — informing and educating health-care professionals and patients or caregivers on the important safety information associated with the use of eculizumab with an emphasis on meningococcal infection *(N. meningitidis)*, other serious infections, and possible serious hemolysis postdiscontinuation.

The approved REMS consists of three main components. First, communication to patients is through a Medication Guide. The REMS also specifies that the health-care professionals who prescribe eculizumab are specially certified and enrolled. The prescribing health-care professional must attest to the following in order to become certified: (i) counsel patients and provide the patient with educational materials including the eculizumab patient safety card and the Medication Guide; (ii) provide the Medication Guide to the patient prior to each infusion; (iii) review the educational materials and the product labeling and comply with the directions for safe use including ensuring meningococcal vaccination status; (iv)

monitor the patient following the infusion for signs and symptoms of serious infections; and (v) promptly report cases of meningococcal infection and other serious adverse events of interest to the company or FDA. Eculizumab will only be distributed to enrolled and certified health-care professionals. The REMS does not directly link documentation of meningococcal vaccination status to administering eculizumab to the patient.

For eculizumab, assessments were required every 6 months from the date of approval of the REMS for 1 year, and annually thereafter. The eculizumab REMS Assessment Plan comprises a number of components including a plan to assess prescriber enrollment/ certification status and discontinuation statistics, an assessment of eculizumab use data, an analysis of cases of meningococcal infection and other serious adverse events of interest, the extent to which the Medication Guide is distributed, an assessment of patients'/caregivers' and health care-providers' understanding about the risks and safe use conditions of eculizumab, and what is known about patients receiving eculizumab, including the number of patients receiving eculizumab, the number of patients who discontinued treatment with eculizumab, a summary of the reasons for discontinuation, the number of patients enrolled in the OneSource Safety Support Program (a voluntary patient support program implemented by the company), the number of patient person-years for enrolled patients on eculizumab, the number of new patients enrolled during the reporting period, the number of patients who received eculizumab who were not enrolled (during the reporting period and cumulative), and the number of patients who were lost to follow up (during the reporting period and cumulatively).

The eculizumab REMS has some interesting features. Because PNH is a rare disease, affecting about 8000 to 10 000 in North American and Western Europe, and a small number of health-care professionals treating them, delivery of safety messages and education about the safe use of eculizumab, including the need for vaccination against *N. meningitides* and to closely monitor for signs and symptoms of meningitis, is relatively simple when compared to products that are used in many patients and by a wide variety of medical specialists.

Meningococcal vaccine is effective against A, C, Y, and W-135 subtypes of meningococcus but are not effective against subtype B, which causes about one-third of meningococcus cases in the US.[24] This underscores the need for close vigilance by patients and caregivers about the need to monitor signs and symptoms of meningococcal infection. As part of the REMS, the patient is to carry a wallet card and present that to the health-care professionals should he or she develop any of the following symptoms and seek treatment:

- headache with nausea and/or vomiting
- headache with fever
- headache with stiff neck or back
- fever of 103°F or higher
- fever and rash
- confusion
- severe muscle aches with flu-like symptoms
- sensitivity to light.

Both REMS programs provide safe access for patients to drugs that are shown to be effective but have serious risks and would otherwise be unavailable.

Risk management approach for eculizumab in the EU

Eculizumab was authorized in the EU in June 2007. Its indication is for "the treatment of patients with paroxysmal nocturnal hemoglobinuria. Evidence of clinical benefit of eculizumab in the treatment of patients with PNH is limited to patients with a history of transfusions." Because of the rareness of the disease (13 cases per million population), eculizumab has an orphan designation in Europe. At the time of authorization, the safety database comprised 195 patients.

The MAH submitted an EU-RMP which included a risk minimization plan. From the safety specification, 11 safety concerns were highlighted: two important identified risks, five potential risks, and four relating to missing information about use in special populations. These were:

- meningococcal infection
- headache
- general infections

- serious hemolysis after drug discontinuation
- infusion reactions
- immunogenicity
- malignancies and hematological abnormalities
- pregnancy and lactation
- children
- renal impairment
- hepatic impairment.

The CHMP decided that there was a need for both additional pharmacovigilance activities and risk minimization activities.

Pharmacovigilance activities. Apart from routine pharmacovigilance, the MAH agreed to institute a global safety registry to obtain information on the most important identified risks and long-term safety data.

Risk minimization activities. There were a number of routine and additional risk minimization activities, which are summarized in Table 29.4. Eculizumab was given a legal status of being subject to restricted medical prescription with administration limited to health-care professionals under the supervision of a physician experienced in the management of patients with hematological disorders. Patients were required to be vaccinated against meningococcal infection at least 2 weeks prior to receiving eculizumab and were required to be revaccinated according to the current guidelines on vaccination. Proof of vaccination was required before treatment started.

The main additional risk minimization activities which were a condition of the marketing authorization were:

1 The MAH shall agree to the details of a distribution system by the National Competent Authorities and must implement such program nationally to ensure that:

- drug distribution will only be possible after checking that the patient has effectively received a meningococcal vaccination with a written confirmation;
- prior to administration, all health-care professionals are provided with information on the following key safety concerns:
  - — headache
  - — infusion reaction
  - — *Neisseria* and general infection
  - — risk of serious hemolysis following eculizumab discontinuation and proposed management
  - — pregnancy and need of adequate contraception in women of childbearing potential
  - — immunogenicity
  - — renal and hepatic impairment.

2 Prior to launch, the MAH shall agree on the implementation of a patient card system in each Member State. This patient card will provide details of the signs and symptoms of infection as well as instruction for the patient to immediately seek medical care. The card will also provide information to health-care professionals that the patient is receiving eculizumab treatment.

3 The MAH will assess the compliance of the prescribers with the recommended Risk Management Plan (RMP) tool by examination of data collected in the PNH Registry. The effectiveness of the risk minimization measures will therefore be assessed by actual clinical practice. Review and assessment of the information collected will be presented in each Periodic Safety Update Report (PSUR).

The Member States were required to ensure that these conditions were implemented in their territories.

During the clinical trial program there were three cases of meningococcal infection of which two were in previously vaccinated patients and one in an unvaccinated patient being treated for idiopathic membranous glomerulonephritis rather than PNH. The course of the illness was different between vaccinated and unvaccinated, with the former having less serious consequences. As a result of these cases it was clear that different risk minimization activities were needed: one to have all patients receiving eculizumab vaccinated prior to starting treatment, and the other in the form of educational material to warn physicians and patients of the risk, signs, and symptoms of meningococcal infection and the need to seek treatment. A patient safety card was also provided. As well as being a reminder to the patient of the risks of meningococcus and other infections, it would also inform any treating physician that the patient was

Table 29.4 The EU risk management plan for eculizumab

Meningococcal infection	**1.** Routine pharmacovigilance **2.** Soliris safety registry: annual survey, maintained at least 5 years, includes collecting information for specific events **3.** Specific reporting including Events of Interest as part of additional pharmacovigilance	**1.** Summary of Product Characteristics (SmPC): Contraindication: patients with unresolved *Neisseria meningitides* infection, not vaccinated patients, patients with known or suspected hereditary complement deficiencies Special warnings and precautions section 4.4: all patients must be vaccinated and re-vaccinated; consideration on appropriate use of antibacterial agents; monitoring, evaluation, and treatment of infections mentioned as the most serious adverse event in section 4.8 **2.** Package leaflet: All patients must be vaccinated against meningococcal infection Vigilance for risks of meningococcal infection Early detection of symptoms of meningococcal infection and steps to manage **3.** Patient safety card: Warning for early detection of symptoms and advice to contact medical facility To be shown to consulted physician for acknowledgement of the risk **4.** Vaccination reminders **5.** Annual physician and patient surveys to assess their understanding of risks **6.** Process to confirm patient *Neisseria* vaccination prior to treatment in each country **7.** Physician's guide **8.** Educational brochure
Headache	1.Routine pharmacovigilance	Mention in section 4.8 of the SmPC and health-care information
General infections	**1.** Routine Pharmacovigilance **2.** Safety registry: Annual survey Maintained at least 5 years Includes collecting information for specific events **3.** Specific reporting including Events of Interest as part of additional pharmacovigilance	**1.** SmPC: Contraindication : patients with known or suspected hereditary complement deficiencies Warning section 4.4 Mentioned as adverse events in section 4.8 **2.** Package leaflet: Vigilance for risks of infections Early detection of symptoms of serious infection and steps to manage **3.** Patient safety card Warning for early detection of symptoms and advice to contact medical facility To be shown to consulted physician for acknowledgement of the risk **4.** Annual physician and patient surveys to assess their understanding of risks **5.** Physician's guide **6.** Educational brochure

Table 29.4 (*Continued*)

Serious hemolysis after drug discontinuation	As above	**1.** SmPC Warning: treatment discontinuation and laboratory monitoring sections in section 4.4 **2.** Package leaflet Vigilance for risks of discontinuation Need to carefully monitor for signs and symptoms of serious hemolysis following drug discontinuation **3.** Annual physician and patient surveys to assess their understanding of risks **4.** Physician's guide **5.** Educational brochure
Infusion reactions	As above	**1.** SmPC: warning in section 4.4 **2.** Package leaflet: sections 2 and 3 **3.** Physician's brochure **4.** Educational brochure
Immunogenicity	As above	**1.** SPC: warning in section 4.4 **2.** Package leaflet : section 2 **3.** Physician's guide
Malignancies, hematologic abnormalities	As above	This risk does not require further mitigation activities
Missing information on use in pregnant and lactating women, children, and patients with either renal or hepatic impairment	Routine pharmacovigilance Prespecified checklist in paroxysmal nocturnal hemoglobinuria safety registry Pharmacokinetic substudy within safety registry	The lack of experience in these special populations is mentioned in Sections 4.2 and 5.2 of the SmPC In section 4.6 of the SmPC: for eculizumab, no clinical data on exposed pregnancies are available Animal reproduction studies have not been conducted with eculizumab (see section 5.3) Human IgG are known to cross human placental barrier, and thus eculizumab may potentially cause terminal complement inhibition in the fetal circulation; therefore, Soliris should be given to a pregnant woman only if clearly needed Woman of childbearing potential have to use adequate contraception methods during treatment and up to 5 months after treatment Advice on contraception is also included in the package leaflet

being treated with eculizumab and so was at increased risk of these infections. Utilizing a controlled distribution ensured that physicians received the educational material prior to eculizumab being available and that a vaccination program was in place.

The future

Managing the risk of medical products, and pharmaceutical products in particular, is an evolving area involving multiple stakeholders in the complex medication use system. Because it is an evolving

area, there are many opportunities for refinements of the current systems that are in place.

One critical area for future development is to measure the impact of risk management strategies. Measurement of this impact is important, because it allows policy makers and other stakeholders to determine if the goals of the strategy are being met. 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, risk management strategies need to be reconsidered. The optimal way to measure the impact of risk management strategies is not well developed, and is an area in which pharmacoepidemiology, along with other disciplines, will play an important role. Pharmacoepidemiology will be integral to the measurement of specific health outcomes, while other disciplines, including social, behavioral, and systems sciences, will be involved to assess the relationship of particular strategies to the 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. A further challenge is to assess if there are negative consequences of risk management strategies. For example, if a restricted distribution is put into place for a certain medicine, does this strategy, because of the real or perceived burdens associated with its implementation, result in patients not receiving the medicine even if the benefits of the medicine outweigh its risks for these patients? If this is the case, might these patients be prescribed other medicines that are not appropriate for them, but which also are not associated with the real or perceived burdens of a restricted distribution system? Questions such as this one are difficult to answer, and will require the integration of many disciplines, including pharmacoepidemiology.

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. Some aspects of risk management, such as providing information to patients, are already incorporated, at least to some degree, in many medication use systems. Other aspects, such as specific risk mitigation measures, are not easily incorporated into current medication use systems. If these risk mitigation strategies are to be used more widely than they are today, or if they are to be used for medicines that are widely used, it will be imperative that systems be developed that integrate into the medication use system.

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 health-care 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 into place across several countries. Second, because of the above differences in the risk mitigation systems that can be put into place across countries, it would be challenging to develop a common risk management document that manufacturers could submit to all regulatory authorities. The differences in risk management activities across countries and regions, however, creates 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*, 2005.
2. Directive 2001/83 EC, Article 1 (28b).
3. National Coordinating Council for Medication Error Reporting and Prevention. http://www.nccmerp.org/aboutMedErrors.html (accessed July 2011).
4. 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.
5. US Food and Drug Administration. *PDUFA Pilot Project. Proprietary Name Review Concept Paper*,

September 2008. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM072229.pdf (accessed July 2011).
6. 21 CFR 201.10
7. Blossom DB, Kallen AJ, Patel PR, Elward A, Robinson L, Gao G, *et al*. Outbreak of adverse reactions associated with contaminated heparin. *N Engl J Med* 2008; **359**: 2674–84.
8. McMahon AW, Pratt RG, Hammad TA, Kozlowski S, Zhou E, Lu S, *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.
9. United States Department Of Health And Human Services. Food and Drug Administration. *Public Workshop Maximizing the Public Health Benefit Of Adverse Event Collection Throughout the Productive Life Cycle*, 2008, pp. 46–64. http://www.fda.gov/ohrms/dockets/dockets/07n0480/FDA-2007-N-0000-TR-(07N-0480).pdf (accessed July 2011).
10. Larson AM, Polson J, Fontan RJ, Davern TJ, Lalani E, Hynan LS, *et al*., and the Acute Liver Failure Study Group. Acetaminophen-induced acute liver failure: Results of a United States multicenter, prospective study. *Hepatology* 2005; **42**: 1364–72.
11. 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.
12. 21 CFR 208.
13. Winterstein AG, Linden S, Lee AE, Fernandez EM, Kimberlin CL. Evaluation of consumer medication information dispensed n retailpharmacies. *Arch Intern Med* 2010; **170**: 1317–24.
14. US Department of Health and Human Services. Food and Drug Administration. *Guidance. Drug Safety Information—FDA's Communication to the Public*, 2007. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM072281.pdf (accessed July 2011).
15. 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. http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/PeripheralandCentralNervousSystemDrugsAdvisoryCommittee/UCM153780.pdf (accessed July 2011).
16. National Asthma Education and Prevention Program's (NAEPP). *Expert Panel Report 3 (EPR-3)*. http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.htm (accessed July 2011).
17. Friedman H, Wilcox T, Reardon G, Crespi S, Yawn BP. A retrospective study of the use of fluticasone propionate/ salmeterol combination as initial asthma controller therapy in a commercially insured population. *Clin Ther* 2008; **30**: 1908–17.
18. 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.
19. 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.
20. Brinker A, Kornegay C, Nourjah P. Trends in adherence to a revised risk management program designed to decrease or eliminaite isotretinoin-exposed pregnancies: Evaluation of the Accutane SMART program. *Arch Dermatol* 2005; **141**: 563–9.
21. 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 Pharmacother* 2007; **41**: 783–801.
22. Shrank WH, Gleason PP, Canning C, Walters C, Heaton AH, Jan S, *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.
23. *Identification of Drug and Biological Products Deemed to Have Risk Evaluation and Mitigation Strategies for Purposes of the Food and Drug Administration Amendments Act of 2007*. 73 Federal Register 60 (27 March 2008), pp. 16313–14.
24. Prevention and control of meningococcal disease: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep* 2005; **54**: 1–21.

CHAPTER 30

FDA's Sentinel Initiative: Active Surveillance to Identify Safety Signals

Judith A. Racoosin, Melissa A. Robb, Rachel E. Sherman, and Janet Woodcock
Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA

Introduction

In May 2008, the Department of Health and Human Services (DHHS) announced the launch of the Food and Drug Administration's (FDA's) Sentinel Initiative to create the national electronic system for medical product safety surveillance mandated by Congress in the FDA Amendments Act of 2007 (FDAAA-PL 110-85).[1] Development of an active surveillance capability through FDA's Sentinel System will give FDA an additional tool for understanding the safety of medical products, and, ideally, create a national resource to investigate their performance. FDA envisions that an enhanced understanding of medical product safety, obtained in a shortened time frame, will result in better information for health-care practitioners, patients, and other consumers, promoting more informed health-care decisions and safer use of medical products. This chapter will describe the goals of the Sentinel Initiative as well as the methodologic and operational challenges that must be surmounted to reach these goals.

FDA's responsibilities in drug safety

FDA is responsible for protecting and promoting health by ensuring the safety, efficacy, and security of human drugs, biological products, and medical devices as well as other FDA-regulated products. FDA is also responsible for helping ensure that the public has access to accurate, comprehensible, science-based information for optimal use of medical products. Postmarketing safety surveillance—monitoring the safety of medical products once they reach the market—is a key component in this effort (see Chapters 8, 10, 19, 20, and 21). For decades, FDA has relied primarily on spontaneous reporting systems to monitor postmarketing safety (see Chapter 10). These systems depend on the public—both health-care practitioners and their patients—to *voluntarily* report adverse events, errors, and quality problems they observe during use to either manufacturers or FDA (medical product manufacturers are legally required to submit to FDA adverse events reports they receive). Each year, FDA receives through its MedWatch and other reporting programs as many as 400 000 adverse event reports on marketed drugs, biological products, and medical devices. Yet, it is broadly recognized that there is substantial under-reporting of adverse events[2] (meaning, many adverse events that are related to medical product use are never reported to FDA; see Chapter 10). The information FDA does receive is highly variable—some reports represent real problems with products, others do not, and many submitted adverse event reports do not always include all the relevant information needed to evaluate a given safety concern. FDA was not alone in recognizing the limitations of this approach and has long been involved in efforts to strengthen postmarketing safety monitoring. Developing an additional medical product safety monitoring capacity that

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

emphasizes what is often referred to as "active surveillance" is a goal that has received increasing attention during the past decade.[3,4]

With the growth in both scope and sophistication of data systems and information technologies and advances in the science of safety—our growing understanding of disease and how medicines work—it was becoming increasingly practicable to explore the creation of a collaborative framework for active surveillance. A critical milestone in this effort occurred in the fall of 2007, when Congress passed FDAAA, in which Section 905 called for active postmarketing drug safety surveillance and analysis.[1]

From the start, the Sentinel Initiative was envisioned as a broad collaborative effort that would explore how best to develop and implement a national electronic system, the Sentinel System, for monitoring FDA-regulated product safety. This system would leverage existing automated healthcare data, including administrative claims data (see Part IIIB), electronic health record (EHR) data (see Chapters 15 and 16), and registries (see Chapter 21), collected for other purposes (i.e., reimbursement, clinical care, and quality evaluations) and often referred to as "secondary use." Once implemented, the Sentinel System would augment FDA's existing postmarketing safety surveillance systems by giving FDA a tool to actively gather and evaluate information about the postmarketing safety and performance of its regulated products. This system is being developed and implemented over time, addressing identified challenges, leveraging existing and emerging technologies, and responding to changes in the evolving landscape of medical informatics. The system will undoubtedly continue to evolve over decades as both the technical capacity and technical methods evolve.

The Sentinel System: implementing active surveillance

FDA has defined a *safety signal* as a concern about an excess of adverse events compared to what is expected to be associated with a product's use.[5] (For purposes of this discussion, adverse event means any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related.) As currently envisioned, the aim of the Sentinel System would be to enhance FDA's available methods for identifying and evaluating postmarketing safety signals in the following ways:

- improving FDA's capability to identify and evaluate safety issues in near real time; and
- enhancing FDA's ability to evaluate safety issues not easily evaluated with the passive surveillance systems currently in place:
  - — expanding FDA's access to subgroups and special populations (e.g., the elderly);
  - — expanding FDA's access to longer-term data; and
  - — expanding FDA's access to adverse events occurring commonly in the general population (e.g., myocardial infarction, fracture) that tend not to get reported to FDA through its passive reporting systems.

Active surveillance is carried out via a continuous, defined process in a specific population and can be conducted using a variety of approaches.[6] Active surveillance can be: medical product-based, identifying adverse events in patients taking certain products; setting-based, identifying adverse events in certain health-care 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). The initial approach being tested in Sentinel Initiative pilots combines medical product- and event-based approaches.

Active surveillance of postmarketing medical product safety typically involves a series of steps, including signal generation and signal refinement. (The terminology used in the fields of pharmacovigilance and pharmacoepidemiology is not well standardized; our discussion here is to provide context for how we are using these terms in our development of active surveillance for the Sentinel System.) *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. Depending on how and when during a product's life cycle a safety signal is identified, a number of approaches for signal refinement can be pursued. Sometimes a safety signal is identified during the product's premarketing review process. In such a case, it might be further refined by monitoring and evaluation during the postmarketing period in a defined population. In other situations, a safety signal may emerge *de novo* in the postmarketing period. When this occurs, the first step is to re-examine data from the clinical development program and consider other available data to look for additional evidence related to the safety signal. A second step may be to conduct a streamlined epidemiologic cohort evaluation to look for evidence of the strength of the association in additional populations.

Should the signal refinement evaluation provide evidence supporting an association between the product exposure and the adverse outcome, it is necessary to conduct additional analyses to validate the signal. The validation process is conducted to ensure that the medical product adverse outcome relationship is not spurious; this process will also likely provide more information about the safety signal, particularly if source record verification is included.

Ultimately, the information gained from a surveillance evaluation in the Sentinel System would be considered along with all other data about the safety signal coming from a variety of other sources, including the premarketing development program (e.g., preclinical, clinical, clinical pharmacology, engineering data), spontaneous reports, and other postmarketing studies to help FDA staff make a regulatory decision. In other words, the evaluation in the Sentinel System will be another tool to provide information to FDA in support of regulatory decision making. Sentinel will help FDA provide useful information to patients and healthcare practitioners to inform decisions and the safe and effective use of medical products.

Investigations informing foundational aspects of the Sentinel System

The launch of the Sentinel Initiative and the subsequent efforts to begin developing the Sentinel System have raised a number of administrative, organizational, and procedural challenges. From the start, FDA has encouraged the participation of all interested stakeholders, recognizing that success would depend on the ability to engage national expertise and secure the commitment of all interested parties. As FDA has worked to develop the Sentinel System, its guiding principles have included:

Integrity: The management structure and data analysis components of Sentinel must be insulated from undue influence.

Privacy protection and data security: It is a fundamental part of FDA's ongoing responsibilities as FDA fulfills its mission to protect public health, including under Sentinel, to safeguard the privacy and security of directly identifiable data and all information FDA receives.

Systems approach: Effective life-cycle safety surveillance of medical products requires a systems approach, including participation of all stakeholders: patients, health-care practitioners, regulated industry, academia, information management, risk communication disciplines, other governmental public health and regulatory agencies, the public, and its elected and appointed representatives.

Transparency: Sentinel's governance structure and processes should incorporate a broad range of expertise and experience as well as address both apparent and actual conflicts of interest.

Consistent with these principles, FDA is relying on a combination of public and expert meetings as well as targeted contracts to execute major portions of the Sentinel Initiative. FDA is making reports on Sentinel activities available to the public on FDA's Sentinel Initiative Web page and in the Agency docket. Contracted research has explored and reported on a variety of topics, including governance and operations, engaging the public, database models, and data sources, among other topics.

A key finding from this work was that a distributed data system is the preferred approach for organizing the active surveillance system. A distributed system allows for data to be maintained in its local environment, as opposed to a centralized approach which consolidates the data into one

physical location. The benefits of this distributed approach include the maintenance of patient privacy by keeping directly identifiable patient information behind local firewalls in its existing protected environment. Because of the data partner's awareness of the changes that have occurred in their health-care system that result in the unique character of each database, use of a distributed system enables the data partner's involvement in running analyses and ensures an informed approach to interpreting results.

Building on the learning to date, in 2009, FDA took the first steps to establish a Sentinel Coordinating Center and competed and awarded a contract to Harvard Pilgrim Health Care Institute. (HPHC) to pilot a miniature Sentinel System (Mini-Sentinel). As part of the pilot, HPHC will use a distributed system wherein a consortium of automated health-care databases is created to build on experience being gained from other ongoing active surveillance efforts and apply the information learned from early Sentinel-related contracts. A goal of the Mini-Sentinel pilot is to create a kind of laboratory that gives FDA the opportunity to test epidemiological and statistical methodologies in the evaluation of postmarketing safety issues and learn more about some of the barriers and challenges, both internal and external, to establishing a Sentinel System for medical product safety monitoring. The collaborating institutions in the consortium include the following organizations: Aetna; Cincinnati Children's Hospital Medical Center; Brigham and Women's Hospital; Duke University School of Medicine; HMO Research Network (includes Group Health Cooperative, Harvard Pilgrim Health Care Institute, HealthPartners Research Foundation, Henry Ford Health System, Lovelace Clinic Foundation, Marshfield Clinic Research Foundation, Meyers Primary Care Institute); HealthCore Inc; Humana; Kaiser Permanente Center for Effectiveness and Safety Research; Outcome Sciences, Inc; University of Illinois at Chicago; University of Iowa, College of Public Health; University of Pennsylvania School of Medicine; Vanderbilt University School of Medicine; Weill Cornell Medical College.

Another important goal of the Mini-Sentinel is to pilot the Coordinating Center model that will enable FDA to submit queries for evaluation to participating data partners who hold automated health-care data containing US patient-level health data that are regularly updated. The data will be maintained within each participating data partner's local, secure environment. Each database will have the capability to link each patient to relevant medical care data, including enrollment status, medical product exposure data, coded medical procedures, and health outcomes. The data sources are to include, to the extent possible, a broad range of patient populations with regard to demographics, socioeconomic status, and co-morbidities.

The Mini-Sentinel Coordinating Center (MSCC) and participating data partners will design and use a common data model (CDM) as the basis for their analytic approach. This approach requires data holders to transform their data into a standardized format. With all data in a common format, the MSCC will be able to write analytic software for a given query, and each participating data partner will run the query in their standardized dataset. Participating organizations will conduct any analyses behind existing firewalls and send only summaries of their results to the MSCC, which will perform further evaluation and compile findings to send to FDA. The use of a common analytic program will minimize the potential for differences in results across data partners caused by use of different analytic approaches.

Importantly, HPHC is assuming responsibility for ensuring that data use complies with the Health Insurance Portability and Accountability Act (HIPAA, see Chapter 35). FDA is not asking for patient-, health-care practitioner-, or health-plan-specific identifiers. Any results provided to FDA are to be aggregated in a standard, predetermined format. As currently envisioned, all analyses of data are to be performed by the data partners in their secure environments without transfer of directly identifiable data.

Many of the issues around active surveillance, described in more detail below, are to be explored in the context of this pilot which will take place over a 3 to 4-year period and, together with information from other ongoing initiatives both in the

United States and abroad, will inform the eventual Sentinel System and may contribute to broadening the use of a distributed system from safety surveillance to other activities, such as comparative effectiveness research (see Chapter 32) or quality assurance (see Chapter 45).

Clinical problems to be addressed by pharmacoepidemiologic research

Marketed medical products are required by federal law to be safe and effective for their intended use in the intended population. A *safe* medical product is one that has reasonable risks, given the magnitude of the benefit expected and the alternatives available. Despite the rigorous US drug development and approval process (see Chapters 1 and 8), well-conducted, randomized, controlled clinical trials cannot uncover every medical product-related safety problem, nor are they expected to do so. In most cases, clinical trials aren't large enough, diverse enough, or long enough in duration to provide all the information on a product's performance and safety (see Chapter 4). Clinical trials are unlikely to detect reliably rare, serious adverse events that would not be expected to occur in a population the size of a premarketing development program, nor would clinical trials identify adverse events with long latency periods or in subpopulations who have not participated in studies. Furthermore, as new medical products enter the market, the potential for interactions with other drugs, biologics, medical devices, and foods increases. Evaluating and updating the evolving safety profile of a medical product, as larger numbers of more diverse patients are exposed during marketing, is a substantive clinical problem. FDA is tasked with providing patients and providers up-to-date information to inform the safe use of medical products. Recent additions to FDA's authorities in the area of drug safety are enabling the agency to meet this need.

New FDA safety authorities

In 2007, with the passage of FDAAA, Congress gave FDA additional authorities and resources to support postmarketing drug safety. For example, FDA now has the authority to require sponsors to conduct postmarketing studies and clinical trials to do the following:[1]

- assess a known serious risk related to the use of a drug;
- assess signals of serious risk related to the use of a drug; and
- identify an unexpected serious risk when available data indicate the potential for a serious risk related to use of a drug.

Additionally, FDA may require sponsors to make safety labeling changes if FDA becomes aware of new safety information that should be included in a drug's labeling. New safety information is defined in Section 901 of FDAAA as "Information derived from a clinical trial, an adverse event report, a postapproval study, or peer-reviewed biomedical literature; data derived from the postmarket risk identification and analysis system…or other scientific data deemed appropriate by the Secretary about (A) a serious risk or an unexpected serious risk associated with use of the drug that the Secretary has become aware of (that may be based on a new analysis of existing information) since the drug was approved, since the REMS was required, or since the last assessment of the approved REMS; or (B) the effectiveness of the approved REMS obtained since the last assessment of the strategy." Finally, FDA can require a sponsor to develop and comply with risk evaluation and mitigation strategies (REMS), when a REMS is necessary to ensure that the benefits of a drug outweigh its risks (see Chapter 43).

There are numerous ways in which the current Sentinel pilot programs, and eventually the Sentinel System, can promote and support the use of these new postmarketing authorities to improve the understanding of the benefit–risk balance of medical products. Active surveillance evaluations in a distributed system could be conducted to determine the level of evidence for new safety information (defined above) that emerges in the postmarketing period. In other cases, active surveillance evaluations could be used to inform the extent to which a REMS is having its desired effect. For example, use of a specific drug could be moni-

tored in near real-time to assess whether it is being used beyond its intended population. Additionally, active surveillance evaluations could be conducted to determine how safeguards put in place via a REMS are affecting a particular clinical outcome.

Emergence of the science of safety

Over the past decade, our understanding of disease origins at the molecular level and our increasing knowledge about the role genes play in how drugs are metabolized are expanding our understanding of how and why medical products cause unintended effects (see Chapter 34). Advances in statistical and epidemiologic methods for active surveillance, combined with advances in medical informatics, are enabling researchers to generate and confirm hypotheses about the existence and causes of safety signals with specific products in defined populations. FDA has begun applying these new techniques to its process for monitoring medical product performance and safety, beginning when a medical product is first being developed on through the application and marketing phases. This life-cycle approach allows safety signals generated at any point during development or marketing to be considered within benefit–risk considerations to inform regulatory decisions and patient care. FDA regards improving the quantification of the benefit–risk analysis to be one of the important facets of the science of safety that urgently requires additional development.

The science of safety also offers new opportunities for addressing a fundamental dilemma: the trade off between safety and access. A clear example of this occurs when FDA, after analysis of a safety signal, considers whether or not to withdraw a drug from the market for safety reasons. Although withdrawal eliminates the possibility of further adverse events, it also deprives those patients for whom the drug is effective. If, using methods being developed today, we can determine that an adverse event is restricted to a small, identifiable segment of the population, the drug, biologic, or device could potentially remain on the market and continue to benefit those who are not subject to the event. There are already some examples of observational data being useful in identifying subpopulations of patients who are at particular risk for a drug-related side effect. In nine trials conducted during the development program for abacavir, a nucleoside analogue indicated in combination with other antiretroviral agents for the treatment of HIV-1 infection, about 8% of patients developed a multiorgan hypersensitivity reaction to the drug (see the Ziagen (abacavir) package labeling, 12/19/2008).[7] Some of these hypersensitivity reactions had severe clinical manifestations, including death. A case–control study[8] and a cohort study[9] identified that HLA allele B*5701 was more common in patients experiencing abacavir-associated hypersensitivity. A subsequent trial using prescreening for HLA B*5701 demonstrated a decrease in risk for abacavir-associated hypersensitivity reactions.[10]

The earliest version of the Sentinel System will likely not have pharmacogenomic data since the incorporation of these data into routine clinical care has been slow;[11] however, the intent of this example is to point out that Sentinel System evaluations may help to identify subsets of patients who are at a differential risk for particular adverse events. Identification of such subgroups will help FDA identify the particular safety signals that would benefit from full pharmacoepidemiologic evaluations, in which particular risk factors for adverse events could be explored.

Methodologic problems to be addressed by pharmacoepidemiologic research

The complex needs of the eventual Sentinel System and the collaborative efforts that are informing its development have raised a number of technological, methodological, and operational challenges that must be addressed as the effort proceeds. Particularly significant methodological challenges include: achieving the goal of real-time or near real-time surveillance; setting the goals of an active surveillance evaluation to produce actionable information in a reasonable time frame; validating a potential safety signal; managing uncertainties in interpreting summary measures from multiple

observational data sources; and developing new methods for signal generation. There are also a number of data-related challenges: developing and implementing harmonized data standards; effectively linking data sources to enhance the length of observation for patients in the population; and enabling access to outcomes that might not be captured in the local database (e.g., device registries, National Death Index). Operational challenges that must be addressed involve creating a national dialogue for discussing issues and establishing relevant policies, developing a comprehensive governance approach, and ensuring privacy and security of data.

Factors affecting the ability to conduct near real-time surveillance

The timing of safety evaluations in the postmarketing period will vary by medical product. For some products, the premarketing development database will suggest that certain signals need immediate postmarketing evaluation. For many medical products, evaluations will also be needed at a later point during product marketing when an unexpected safety signal emerges.

As discussed in the following sections, the ability to conduct near real-time surveillance using the Sentinel System will be affected by how quickly a medical product is *taken up* into the marketplace as well as how often data sources are *updated* with new exposures and outcomes of interest as they occur.

Uptake of new products into the market

Uptake of a new medical product into the market is influenced by a variety of factors. For patients who are part of a particular health plan, the formulary of that health plan will influence what medical products are used. Evidence-based practice guidelines, promulgated either by professional societies or the medical literature, can also shape product uptake. In each individual case, a variety of these factors will interact to shape the timing of product uptake.

An assessment conducted to inform the development of the Sentinel System about the timing of uptake for two sample pharmaceuticals in two health-care settings illustrates this point.[12] The two drugs, sitagliptin and duloxetine, were selected for their relatively recent approval, as well as their approved indication for common conditions, diabetes and depression, respectively. The uptake of sitagliptin, approved in October 2006, was documented for the period October 2006 to June 2008. The uptake of duloxetine, approved in August 2004, was documented for the period August 2004 to July 2006. The two health-care settings were the following:

- IMS Health Plan Claims Databases (HP), formerly known as Pharmetrics, is a patient-centric database comprising data submitted by multiple commercial insurance companies with about 12 million persons having continuous enrollment for 1 year, 10 million for 2 years.
- MHCD Database (MHCD) represents the population enrolled in TRICARE, the health-care program serving active-duty service members, National Guard and Reserve members, retirees, their families, survivors, and certain former spouses, worldwide. About 8 million people have continuous enrollment over 2 years.

Figure 30.1 shows a comparison of duloxetine uptake in these two health-care settings, and Figure 30.2 shows a comparison of sitagliptin uptake in the same two health-care settings.

The figures illustrate the types of variability FDA will encounter when trying to understand the influence of uptake patterns on active surveillance evaluation results. The differences in uptake between the two products in the two health-care settings could reflect different formulary practices, the position of the new drug among other treatments approved for the indication, physician preference, and/or the demographic make-up of the population covered in the health-care plan.

Another way the authors measured the time of uptake of these medical products was to identify the number of months needed to accumulate a certain number of unique drug exposures. Table 30.1 includes those patients who *ever* received a prescription for the drugs of interest during the study period. As shown below, a threshold of 1000 patients exposed was achieved by month 1. However, there was variability between the two data sources in the time to achieve higher numbers

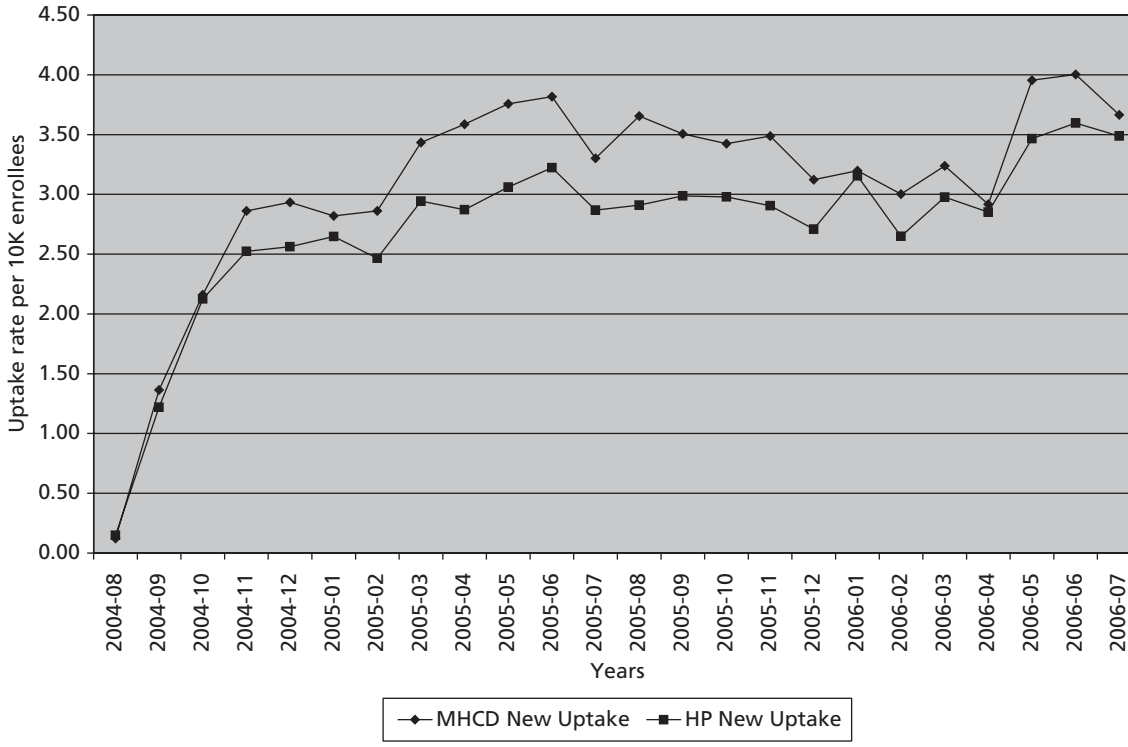


Figure 30.1 Duloxetine uptake in HP and MHCD databases. HP data, IMS Health, IMS Health Lifelink Health Plan Data™, years 2004–2006, data extracted December 2008. MHCD data, DoD, Military Healthcare Data, years 2004–2006, data extracted December 2008. Adapted from IMS Government Solutions[12] with permission.

of patients. Only the commercial database reached the threshold of 100 000 patients exposed to one of the drugs, duloxetine, within the time frame of the study.

Persistence of treatment may be important for assessing adverse events that don't occur in the first month of treatment. The population described in Table 30.2 is limited to those patients who continued to fill prescriptions for the drugs of interest for 6 months or more. When the population is limited in this way, it takes more months to reach the various thresholds of patients exposed. As shown below, none of the populations of patients treated for 6 months or more reached the threshold of 100 000.

These data demonstrate that the requirement for a larger number of treated patients or a group of treated patients who have a certain persistence of therapy will extend the time to reaching a target level of patients exposed. Active surveillance of adverse events that only appear with prolonged drug exposure also will take longer to detect.

Frequency of data updates

Another factor that will affect the ability to achieve near real-time surveillance is the frequency with which data are updated and become available for analysis. To monitor the safety of newly approved therapies optimally, data that are as current as possible will be needed. Some medical products will require more frequent data updates than others. For example, annual influenza vaccination occurs in a relatively short period of time over a series of weeks to a few months. Similarly, newly approved vaccines for other infectious diseases often have widespread rapid uptake in the population. Thus,

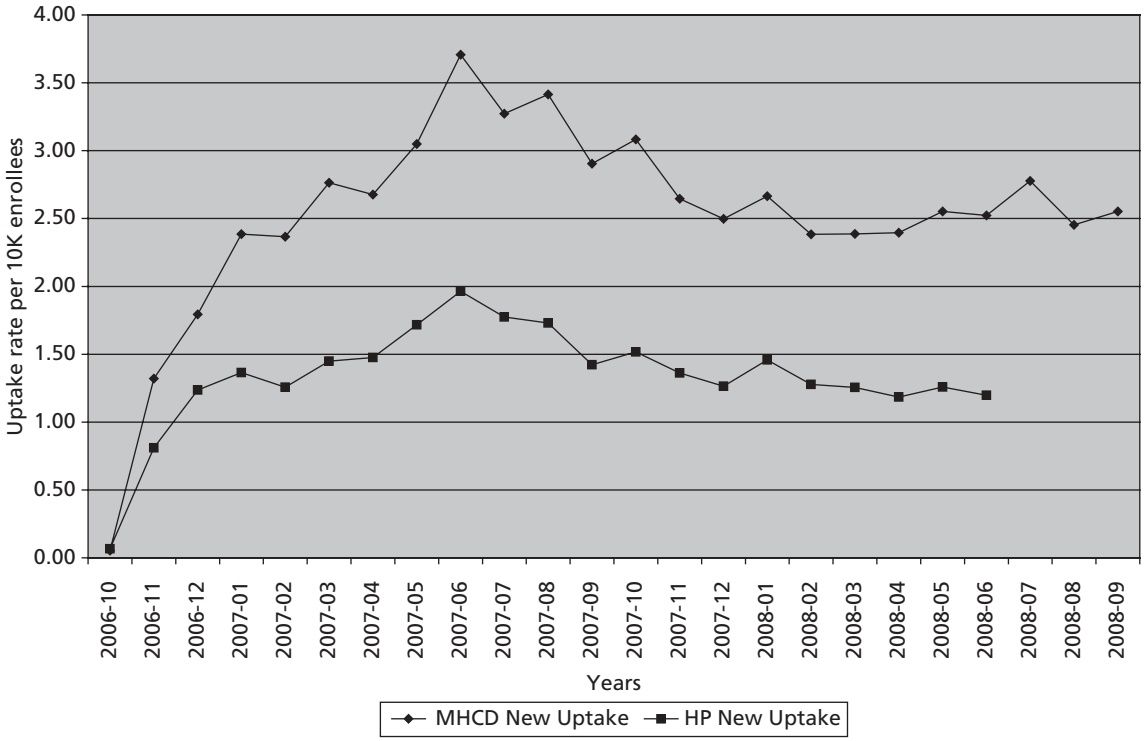


Figure 30.2 Sitagliptin uptake in HP and MHCD databases. HP data, IMS Health, IMS Health Lifelink Health Plan Data™, years 2006–2008, data extracted December 2008. MHCD data, DoD, Military Healthcare Data, years 2006–2008, data extracted December 2008. Adapted from IMS Government Solutions[12] with permission.

Table 30.1 Patients who ***ever*** received a prescription for drugs of interest during study period

Number of patients exposed	Months after approval			
	HP		MHCD	
	Duloxetine	Sitagliptin	Duloxetine	Sitagliptin
100	<1	1	<1	1
1000	1	1	1	1
10000	3	6	5	6
100000	21	N/A	N/A	N/A

HP data, IMS Health, IMS Health Lifelink Health Plan Data™, years 2004–2008, data extracted December 2008.
MHCD data, DoD, Military Healthcare Data, years 2004–2008, data extracted December 2008.

Table 30.2 Those patients remaining on drugs of interest for 6 months or more

Number of patients exposed	Months after approval			
	HP		MHCD	
	Duloxetine	Sitagliptin	Duloxetine	Sitagliptin
100	<1	1	1	1
1000	1	2	2	1
10 000	6	8	8	7
100 000	N/A	N/A	N/A	N/A

HP data, IMS Health, IMS Health Lifelink Health Plan Data™, years 2004–2008, data extracted December 2008.
MHCD data, DoD, Military Healthcare Data, years 2004–2008, data extracted December 2008.

in many cases, efficient monitoring for safety problems with vaccines requires weekly updates on vaccine exposure and adverse outcomes of interest (see also Chapter 26).[13] The Vaccine Safety Datalink (VSD), an active surveillance system[14] for newly approved vaccines run by the Centers for Disease Control and Prevention (CDC), gets weekly updates from eight health maintenance organizations (HMOs) with vaccine exposure and inpatient and outpatient diagnostic information for outcomes of interest (see Chapters 12 and 26).[15] However, early studies that have begun to simulate applying sequential testing methods used in vaccine safety assessment to drug-related safety signals have used monthly intervals to capture drug exposures, recognizing that drug uptake occurs more slowly than vaccine uptake.[16] The frequency for data updates for active surveillance will be driven by the expected speed of product uptake as well as the urgency with which more information about a safety signal is needed. Additionally, a practical consideration that impacts data updates is the frequency at which data partners are able to make data available for evaluation. In the case of administrative claims data (see Part IIIB), many systems take 3 to 6 months to collect, adjudicate, and prepare data for analysis. The adjudication and data preparation processes affect the time frame for data to be available for surveillance purposes.[17]

Validating a safety signal

It is incumbent on the team reviewing the results of a safety signal evaluation to rule out causes of a potential false-positive result.

The VSD team at the CDC and their HMO Research Network-based collaborators have extensive experience vetting safety signals in search of true positives (see Chapter 26). When applying the Maximized Sequential Probability Ratio Test that has been developed for real-time monitoring of vaccine-related adverse outcomes (see Chapter 46), the VSD investigators often use two types of control groups, matched controls and historical expected counts, which have complementary advantages and disadvantages.[16] Recognizing that the claims data submitted on a weekly basis for monitoring become more complete over time, subsequent confirmatory analyses are conducted in some cases later in the evaluation period to take advantage of more mature data. Another technique that has been applied to assess potential biological plausibility of a safety signal is the use of a temporal scan statistic to test for temporal clustering in the postexposure time period. Additionally, logistic regression analyses, adjusting for relevant covariates, are sometimes conducted to better assess the risk of the adverse outcome.[18] Validation of coded diagnoses against source records may also be conducted. If a signal persists, VSD investigators may conduct

additional analyses or design a full study to test the hypotheses raised by the signal.[16]

Many similar issues will need to be considered as safety signals emerge from active surveillance evaluations conducted for drugs, biologics, and devices within the Sentinel System. Evaluations of safety signals for medical products used in sick patients (in contrast to vaccines which are generally given to healthy individuals) will present additional methodologic issues, in particular, confounding by indication (i.e., the sickest patients tend to be started on the newest drugs and pushed to the highest doses, see Chapters 3 and 47).

An additional issue related to validating safety signals that must be grappled with is the question of whether the same database can be used for both signal generation and hypothesis confirmation. One practice is to split a database so that different data are used for signal generation component and the hypothesis confirmation component.[19] However, other researchers advocate the use of an independent data source to test hypotheses.[20] The Observational Medical Outcomes Partnership program (see section on New Methods) is conducting empirical evaluation of each approach to inform this controversy.

Interpreting summary measures from multiple observational data sources

As described above, in a distributed system, as is envisioned for the Sentinel System and being piloted in Mini-Sentinel, data partners would run evaluations on their data and return only summary measures to a Coordinating Center. Some methods used to evaluate safety signals may generate counts that would then be summed by the Coordinating Center, whereas other methods would generate a summary estimate at each individual site.

How to interpret varying estimates of the strength of the safety signal from each individual site remains unclear. The data that will be leveraged by the Sentinel System for these surveillance evaluations are being collected for other purposes (e.g., insurance reimbursement, clinical care), not for medical product safety surveillance. Because each institution has a unique way in which it uses its formularies and applies diagnostic coding practices, variations in the results of a safety signal evaluation may be driven by variations in institutional practices.

Additionally, we must consider whether and how such results may be combined. In this case, although each data partner's evaluation is not a full pharmacoepidemiologic study, we may look to the guidelines on meta-analyses of observational studies[21] to inform our thinking (see Chapter 40). The application of formal meta-analytic techniques to observational studies must be conducted carefully because observational studies are prone to certain confounding and biases that are not an issue with randomized controlled trials.[22] One of the major concerns, heterogeneity of study design, is addressed by the use of a centralized analytic approach run on data that have been standardized in a CDM. However, evaluation of other sources of heterogeneity (e.g., patient population, formulary variation, prescribing practices, approach to diagnostic coding) must be actively pursued. Ultimately, a combined estimate of the strength of the safety signal may be useful, but the consideration of reasons for heterogeneous results across data partners, if it occurs, is likely to be informative as well.

Developing new methods for signal generation

Based on strategic and resource considerations, FDA has elected to develop signal refinement capabilities as the initial focus of the Sentinel Initiative. Ultimately, however, the Initiative will also need to turn its attention to signal generation.

Statistical methods for signal generation are used in many fields—developing spam filters and identifying credit card misuse are two. However, the risk attendant to any signal generation effort with medical product safety is the potential, significant, public health effect of a false-positive signal. In addition to the resources, both financial and personnel, required to discern true positive signals from false positive signals, there are consequences that must be considered when determining when to communicate a safety signal identified using signal generation techniques. Patients who learn that their medical therapy may carry a safety risk

may cease using the therapy. If it later turns out that the safety concern is unsubstantiated, patients may have suffered because they discontinued the needed therapy. Additionally, in this era of increased access to information, we must be mindful of the potential effect on health-care practitioners related to alert fatigue. FDA would like to minimize the risk of desensitizing health-care practitioners and the public to alerts of public health importance by not communicating false-positive safety signals.[23]

Over the past decade, data mining methods for spontaneous reports have been developed for signal generation by those working in the pharmacovigilance field, including FDA, pharmaceutical companies, academia, and vendors (see Chapter 10).[24,25] More recently, these methods are being applied to observational databases.[26] FDA will be looking at research being conducted to maximize the collection of actionable information from signal generation methods while minimizing the occurrence of false positives.

One group that is actively evaluating methods for signal generation in observational data is the Observational Medical Outcomes Partnership (OMOP). OMOP is a public–private partnership managed through the Foundation for the National Institutes of Health and funded primarily through donations from the pharmaceutical industry.[27] OMOP is conducting a 3-year initiative to develop and test methods that are feasible and useful for analyzing existing health-care data to identify and evaluate safety and benefit issues related to drugs already on the market. OMOP's research agenda includes development and assessment of methods for both signal refinement and signal generation. In the area of signal generation, OMOP is evaluating a series of approaches to identifying associations between drugs and conditions for which the relationships were previously known.

OMOP is using an empirical approach for testing the utility of a range of methods for signal generation. No preselection has been made to narrow the types of methods tested for signal generation from the pool also being evaluated for signal refinement activities. Rather, the same group of methods being tested on targeted drug–health outcome of interest (HOI) pairs for signal refinement purposes is being tested on the non-specified outcomes test cases. OMOP is testing various types of methods, including disproportionality analyses, exposure-based methods, and case-based methods. The results of the evaluations of these methods for signal generation are expected in 2011.

Data infrastructure needs to conduct active surveillance

To achieve a modern health information environment, we need to enhance and integrate three key information management domains: (i) *access* to information; (ii) *interfaces,* or user-friendly tools, supported by a robust IT architecture, to convert information efficiently into knowledge; and (iii) *standards* that are used by all to facilitate information exchange. Improving *access* to data sources alone is not enough. We need better *interface tools*, and they cannot work efficiently without *standards*. These three domains interact to influence the efficiency with which we receive, manage, and communicate information.

Data structure

The eventual Sentinel System is planned to be a distributed system with data held in local environments. FDA evaluated the optimal characteristics of a possible database model for such a system. The conclusions[17] were that a CDM be used to facilitate the conduct of active surveillance evaluations, and the following key advantages of creating and employing a CDM for the Sentinel System were identified:

- In the absence of a nationwide, interoperable health information technology infrastructure, based on federally recognized and widely adopted standards, it would create a common language that can be used across data sources for evaluations.
- A CDM would reduce the need centrally to create and maintain complex metadata and data-mapping activities.
- A CDM would address issues related to ensuring that all data partners are using the same terms to describe the same concepts.
- It would enable a centralized analytic approach.
- It would concentrate data-validation efforts on the initial data transformations and not for each project.

However, disadvantages were also identified that should be considered. In addition to substantial start-up effort and expense, additional resources must be invested to perform data updates since near real-time active surveillance will require updating data elements on a periodic schedule as previously explained. More importantly, a CDM may not be able to capture the full granularity of source data.

To evaluate the strengths and weaknesses of the CDM approach, FDA is learning from a series of active surveillance pilots. Mini-Sentinel and OMOP are using the CDM approach. OMOP also has centrally held databases on which they are conducting some assessment of active surveillance evaluations that do not use a CDM. FDA also has a pilot project called the Federal Partners Collaboration (FPC), which is creating a small distributed system with the Centers for Medicare & Medicaid Services, the Department of Defense, and the Veterans Health Administration. The FPC will not be using a CDM but will use a common evaluation protocol. Therefore, this pilot will provide an opportunity to explore the challenges associated with interpreting safety signals that come from a system of databases not using standardized terminology or centralized analytics. The lessons learned from these pilots will inform the data structure for the eventual Sentinel System.

Data standards

The Sentinel System is predicated on the secondary use of existing, automated, health-care data to obtain a better understanding of medical product safety. The lack of a standard format for the storage and exchange of data slows the sharing of information among key stakeholders, making it cumbersome and manually labor intensive. The development and use of terminology standards for important data elements used by the Sentinel System, particularly for electronic health records (EHRs), could greatly facilitate the creation of analytical tools. Standardizing data elements and terminologies is critical to any attempt to achieve a modern electronic approach to monitoring medical product safety.

FDA is collaborating with the Office of the National Coordinator for Health Information Technology (ONC), which is seeking to adopt and harmonize a set of standards and implementation specifications. In an interim final rule issued January 2010,[28] the ONC outlined criteria aimed to directly support objectives for meaningful use of EHR data. Longer-term objectives that may be facilitated by the Sentinel System include: promoting improvements in quality, safety, and efficiency; and improvement of population health. Creating and adopting these standards for health-care IT systems will greatly improve the ability of the Sentinel System to effectively use these EHR data for postmarketing surveillance.

Database linkages

As part of the foundational work that FDA is doing to inform the eventual Sentinel System, the agency must understand what data are collected, the duration of patient observation, and the completeness of data. This is particularly critical because, under our fragmented health-care system, Americans receive health care from a myriad of organizations. It is unrealistic to expect that the Sentinel System will ever realize its full potential without eventually linking information at the patient level, yet there are many technical, privacy, and security obstacles to overcome before this potential can be realized.

Most individuals move across medical systems and insurers, making it difficult to track individuals over time. The ability to link individuals across health insurer systems (as they switch insurance carriers) and EHR systems (as they switch providers) would enable long-term follow-up and remove a substantial limitation of current observational studies using routinely collected health-care information. It would also diminish the risk of patients being double counted because they appear in more than one health-care system or administrative claims database.

Beyond linking across health-care practitioners and systems, there is also the need for the eventual Sentinel System to link to external data sources such as vital statistics databases or registries

(e.g., National Center for Health Statistics, birth and death registries, tumor registries) to capture patient data beyond that found in administrative claims databases or EHRs. This type of linkage is common in epidemiologic research. For example, it is possible to identify deaths from the National Death Index (NDI) and ascertain childbirth information from State birth registries. The HHS Post-licensure Rapid Immunization Safety Monitoring (PRISM) project[29] is an example of an active surveillance system that has linked state immunization registries with health-care plans to monitor adverse events related to the H1N1 vaccine.

Although integration of multiple disparate sources could improve the understanding of each patient as he or she encounters the health-care system, such integration across systems requires surmounting substantial regulatory, privacy, and technical challenges (see also Chapter 35). Linkage across data sources requires the ability to create unique individual identifiers and use them in all relevant systems. Such a unique identifier usually is composed from protected health information (PHI) (e.g., name, date of birth, address, and/or social security number). The Privacy Rule protects all "individually identifiable health information" held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral.. Within a distributed system as envisioned for the Sentinel System, the data partner would have direct access to the needed information for the linkage and would be the most reasonable institution to request such a linkage. In addition, in a distributed system, PHI would only be shared between the two institutions that need the information to conduct the linkage (i.e., health plan and curator of the external data source). Minimizing the number of entities involved in a linkage substantially reduces the potential legal complication of the linkage. Regional health information exchanges, such as Indiana Health Information Exchange[30] and the South Carolina Health Information Exchange,[31] have advanced the informatics technologies needed to facilitate linking data sources to facilitate the movement of patient information within their systems to enhance clinical care and public health.

Operational challenges

In addition to the methodologic challenges noted above, there are numerous operational challenges underlying the development of a system such as the one envisioned. FDA has, with broad public input, begun creating the foundation for what is hoped to be a valuable national resource to help improve the informed, safe use of medical products.

National dialogue

FDA has developed, and is continuing, a national dialogue with experts in relevant disciplines (e.g., pharmacoepidemiology, statistics, informatics, and privacy). FDA is engaging relevant stakeholder groups, including patients, consumers, health-care practitioners, automated health-care data partners, and regulated industry.

The importance of this approach cannot be overemphasized. First, there is significant knowledge and expertise that FDA must tap into if the Sentinel System is to be optimally developed. Second, because the idea of a distributed data system is a new, formidable concept, FDA must be completely transparent about what it is doing and how it is being done. Third, because this effort is on the leading edge in a variety of areas, FDA must ensure that findings are shared among those who are developing related activities. Finally, FDA is committed to minimizing duplication of efforts, resources, and/or functionality and that findings resulting from this effort are placed in the public domain.

Governance

FDA's thinking on how the Sentinel System should be governed has evolved significantly. FDA remains committed to taking the lead in this effort, yet, as mentioned above, collaboration is critical, and expertise and resources need to be shared to the greatest extent possible—in fact, FDAAA mandates collaboration. FDA has also concluded that its statutory mandates require it to exert substantial control over certain activities. Thus, a portion of

Sentinel operations must remain under agency control. This model is being tested as part of the Mini-Sentinel contract. A major challenge in the years ahead will be to develop a governance framework for the national resource portion that ensures broad participation and reflects a sustainable business model.

Privacy and security

Since the launch of the Sentinel Initiative in 2008, FDA has engaged thought leaders in the privacy and security field. One of the first contracts awarded under the initiative involved identifying and analyzing potential privacy issues (see Chapter 35). We have already described above how a distributed system approach maintains patient privacy by keeping directly identifiable patient information with data partners behind local firewalls.

As we have begun to grapple with the realities of conducting active medical product surveillance, we have come to understand that there may be infrequent occurrences when de-identified datasets may not be sufficient to efficiently monitor the safety of a particular medical product. For example, if it becomes necessary to validate a coded HOI against source records, it may be necessary to access some PHI (protected health information). In some cases, if claims data are used for an evaluation, some elements of PHI (e.g., month, year of an exposure) may need to be transferred from the health-care environment that delivered the care to the claims environment that is conducting the evaluation. Additionally, in the case when a statistical evaluation results in a small number of patients identified in a cell, even though elements of PHI may not be included in the summary results, the summary results themselves may not be considered de-identified due to the small number of patients in the cell. In some cases, the Health Insurance Portability and Accountability Act (HIPAA) requires data use agreements to be established that outline the specific purposes of the data exchange and the procedures to be put in place to guarantee protection of PHI. The Privacy Rule permits covered entities to disclose PHI, without authorization, to public health authorities who are legally authorized to receive such reports for the purpose of preventing or controlling disease, injury, or disability. This would include, for example, the reporting of a disease or injury; reporting vital events, such as births or deaths; and conducting public health surveillance, investigations, or interventions (see 45 CFR 164.512(b)(1)(i)). HIPAA's Privacy Act does not restrict de-identified datasets from either use or disclosure.[32] There are two ways HIPAA makes de-identification of a dataset possible. The first involves a qualified statistician determining and documenting that "the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual."[32] Alternatively, a dataset can be de-identified using the *safe harbor* approach (see Box 30.1).

There may also be instances that require information sharing between data partners and a Coordinating Center, or among data partners, as noted previously. FDA is actively exploring methods and techniques to ensure that only the minimum amount of PHI necessary leaves its local environment to meet the needs of a specific active surveillance evaluation. FDA is also actively engaging with privacy experts to explore what other avenues of consumer protection should be developed or expanded.

Other privacy and security regulations apply to the Sentinel System, similar to existing systems that process, publish, transmit, or store FDA information or information on behalf of FDA. The eventual Sentinel System will have to be protected in accordance with the Federal Information Security Management Act (FISMA), issued in 2002.[33] The E-Government Act (Public Law 107-347), which was passed by the 107th Congress and signed into law by the President in December 2002, recognized the importance of information security to the economic and national security interests of the United States. Title III of the E-Government Act (the Federal Information Security Management Act (FISMA)) requires each federal agency to develop, document, and implement an agency-wide program to provide information security for the information and information systems that support the operations and assets of the agency, including those pro-

Box 30.1 Health Insurance Portability and Accountability Act (HIPAA) de-identification

HIPAA* allows for de-identification of a dataset by removing the following identifiers of the individual or of relatives, employers, or household members of the individual.

1. Names
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly available data from the Bureau of Census
   - a. the geographic units formed by combining all zip codes with the same three initial digits contains more than 20 000 people, and
   - b. the initial three digits of a zip code for all such geographic units containing 20 000 or fewer people is changed to 000
3. All elements of dates (except year) for dates directly related to the individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
4. Telephone numbers
5. Fax numbers
6. Electronic mail addresses
7. Social security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/ license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. Web Universal Resource Locators (URLs)
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code, except as permitted for re-identification purposes provided certain conditions are met.

In addition to the removal of the above-stated identifiers, the covered entity may not have actual knowledge that the remaining information could be used alone or in combination with any other information to identify an individual who is subject of the information.

**45 CFR 164.514(b)(2)*

vided or managed by another agency, contractor, or other source. Because the Sentinel System is being sponsored by FDA and is being established in response to FDAAA, the Sentinel System must be assessed as part of FDA's Certification and Accreditation (C&A) process as required by FISMA. Due to the nature of the types of data being used in Sentinel, there is minimal, if any, risk of security breaches resulting in disclosure of PHI. FDA recognizes, however, that attention will need to be paid to computer security with respect to the transmission of queries and results summaries, and FDA will require implementation of policies and procedures to ensure computer security at each stage of the process.

Currently available solutions

FDA uses a number of programs and tools to carry out postmarketing surveillance, but efforts focus primarily on its spontaneous adverse events reporting system, a mostly passive system comprising the Adverse Event Reporting System (AERS) (see Chapter 10),[34] the Vaccine Adverse Event Reporting System (VAERS) (see Chapter 26),[35] and the Manufacturer and User Device Experience (MAUDE) Database (see Chapter 27).[36] As efforts to develop the Sentinel System continue, FDA is increasingly looking to learn from the active surveillance programs already in place to help inform the development of the Sentinel System.

The Vaccine Safety Datalink (VSD)

One initiative that has used active surveillance methods to monitor medical product safety is the VSD, a collaboration between the CDC's Immunization Safety Office and eight managed care organizations (see Chapter 26).[14] Since 2006, the VSD has conducted rapid-cycle analyses (RCA) to monitor for safety signals related to vaccinations in near real-time, using automated systems that are able to track medical product exposures (specifically, immunizations) and inpatient, emergency department, and outpatient diagnoses for 8.8

million patients. As described on the VSD Website, RCA works in the following way:

> Each week, the rate of adverse events that occurs in people who have received a particular vaccine are compared to the rate of adverse events that occurs in a similar group of people who have not received that vaccine. If the rate of adverse events among vaccinated people is significantly higher than among the comparison group, the vaccine may be associated with an adverse event. To find out if a vaccine truly increases the risk of a particular adverse event, VSD Project scientists conduct a formal epidemiologic study.

The vaccines that have been studied include meningococcal conjugate (MCV4), tetanus–diphtheria–acellular pertussis (Tdap), measles–mumps–rubella–varicella (MMRV), rotavirus, human papillomavirus, and influenza.[37] For each vaccine, five to ten outcomes were selected for monitoring prior to data collection, based on biologic plausibility and safety concerns that emerged during vaccine development programs. Event rates following vaccination were compared to event rates after historical or concurrent control visits using sequential testing methods.[15]

Among the 11 positive signals identified through the RCA approach in the last 5 years, one signal has been considered to represent a true increase in vaccine-associated risk. During the measles–mumps–rubella–varicella (MMRV) vaccine prelicensure study, there was a signal for fever at 5–12 days and 0–42 days postvaccination. Postmarketing monitoring by VSD detected an increased risk of post-MMRV seizure in 12- to 23-month-old children compared to those receiving MMR (measles–mumps–rubella) vaccine. A follow-up pharmacoepidemiologic study compared the postvaccination seizure rate with MMRV to that observed with the separately administered measles–mump–rubella and varicella vaccines (MMR+V); the seizure rate following MMRV was 9/10 000 as compared to a post MMR + V seizure rate of 4/10 000. The adjusted odds ratio was 2.3 (95% CI: 1.6–3.2). The results of this study led to a policy change. When VSD study results were considered along with similar results from a Merck postlicensure study, the CDC's Advisory Committee on Immunization Practices changed its recommendation from MMRV being the preferred choice over MMR + V to no preference between MMRV and MMR + V.[38] This example demonstrates VSD's role in active surveillance for signals emerging from vaccine development programs.

Postlicensure Rapid Immunization Safety Monitoring (PRISM)

The goals of the PRISM[29] system are to assess large populations for rare events that may occur following H1N1 influenza vaccination in as near real-time as possible, including patients who received H1N1 vaccine from both public and private providers. The vaccine exposure and outcome data come from large health plans covering approximately 30 million people, with vaccine exposure data from immunization information systems (IIS) in eight states covering 17 million people. The system includes active surveillance for increased risk of key prespecified conditions including the following: anaphylaxis and other allergic reactions, ataxia, Bell's palsy and other cranial nerve disorders, demyelinating disease, encephalitis, Guillain–Barré syndrome, myocarditis, neuropathies, preeclampsia, seizures, spontaneous abortion, and stillbirth; and a capability to investigate unanticipated specific concerns that may arise based on signals from VAERS or other sources.

The PRISM system benefits from using many components of the well-tested VSD system while capturing a much larger population than VSD. Additionally, PRISM has been designed to link state registry exposure data to health-plan data. (The novel state registry component of the project depends on states being able to capture the majority of public setting vaccinees in a timely and complete fashion.) Two types of control groups are being used in the PRISM system: expected counts based on historical rates and unexposed time periods in the same persons. Variations of the maxSPRT statistical test are being applied to each analysis.

At a meeting of the National Vaccine Advisory Committee, various groups conducting active surveillance on H1N1 influenza vaccine presented interim findings.[39] For PRISM, among the 30 million people surveilled by February 20, 2010, 1.9 million people had received an H1N1 vaccination. Using the VSD's RCA approach described above, no

signals for the prespecified outcomes had been identified.

Early detection of adverse drug events by integrative mining of clinical records and biomedical knowledge (EU-ADR)

The EU-ADR effort is a European active surveillance initiative, launched in February 2008 and funded by the European Commission.[40] The objective of EU-ADR is to "design, develop and validate a computerized system that exploits data from EHRs and biomedical databases for the early detection of adverse drug reactions."[41] The system links a total of 30 million electronic patient records from eight databases in four Member States (UK, Denmark, Netherlands, and Italy). Data are extracted locally on a periodic basis, drugs and events are mapped to a common terminology, and data are pooled for the purposes of signal generation and signal validation.

EU-ADR is using an event-based approach, where a set of specific events are evaluated for their association with all possible drugs. Recognizing that a broad data mining effort may create a lot of false-positive signals, they aimed to identify specific events that would be important to evaluate as part of a pharmacovigilance program. To do this, they ranked adverse events by public health importance on the following criteria: frequency of the event as trigger for drug withdrawal; frequency of the event as trigger for a black box warning; leading to emergency department visit or hospital admission; probability of event to be drug-related; and likelihood of death. Based on the ratings, cutaneous bullous eruptions, acute renal failure, anaphylactic shock, acute myocardial infarction, and rhabdomyolysis were ranked at the top of 23 total events, and will be the initial focus of EU-ADR's signal detection efforts.[41]

In an effort to validate data mining techniques for signal detection, EU-ADR researchers evaluated a known positive association between non-steroidal anti-inflammatory drugs (NSAIDs) and upper GI bleeding. The relative risks of NSAIDs ranged from 2.1 to 5.0, consistent with what has been cited in the literature. Each database was able to find the association between NSAIDs and upper GI bleeding, providing confidence that the data mining techniques were effective at detecting a known positive association.[42]

VSD, PRISM, and EU-ADR have each grappled with data infrastructure, methods, and operational challenges that will face the eventual Sentinel System. FDA is studying these programs closely to benefit from the lessons they have learned along the way and apply those learnings to facilitate the development of the Sentinel System.

The future

The overarching challenge is to ensure that that the Sentinel System it is sufficiently nimble to meet the changing needs of a nation in which the health-care system is rapidly evolving. This means that the Sentinel System must be poised to be integrated into other national efforts aimed at secondary use of health-care data. Within that context, it is important to note that these efforts will not ever optimally serve the public until the nation has learned how to integrate clinical care with clinical research.

Certain challenges are paramount and have been discussed in detail elsewhere in this chapter. Health care is delivered in a myriad of disparate settings; we must learn to collect information, convert it so that it can be analyzed in a standardized manner, and we must learn how to link data while appropriately protecting patient privacy and data security.

Certain product classes have unique problems. For example, the absence of unique identifiers for medical devices makes tracking these products particularly difficult. Similarly, for injectable biological drugs, the dependence on CPT (Current Procedural Terminology) codes can make safety surveillance a challenge. These are examples of infrastructure problems that the nation must address to better protect the public.

Sentinel is a long-term, complex initiative that will, by necessity, be implemented in stages and will evolve as capabilities, methods, public awareness, public acceptance, and data standardization increase. We must balance eagerness for progress

with the very real risks inherent in overreaching current capabilities. Moreover, while recognizing that FDA cannot and should not attempt to develop Sentinel in a vacuum—this is neither wise nor feasible—we are mindful, and must frequently remind others, that as a regulatory agency with statutory mandates, we cannot and will not share those responsibilities and decisions.

As we build the Sentinel System, we must create electronic interfaces that can send queries to existing data sources consistent with appropriate privacy guidelines and applicable laws. The ongoing development and deployment of health-care system-based EHRs for clinical encounters, laboratory, and other diagnostic data, occurring both in hospital and outpatient settings, offer important opportunities to query a variety of sources quickly. To maximize the usefulness of active surveillance with automated health-care data, methods need to be fine-tuned to link information about patients among data sources. We and others must focus resources on ensuring that evaluations are able to provide a complete longitudinal profile of patient care, to ensure greater confidence in safety signals identified and refined within this system.

In addition, more attention needs to focus on creating and adapting existing statistical and epidemiologic methods for use in a distributed surveillance system for a broad spectrum of pharmacovigilance activities, including signal generation, signal refinement, and signal validation (see Chapter 46). As a nation, we must ensure that we train the next generation of experts in this and related fields.

In the years ahead, we must establish a governance framework that permits a portion of the Sentinel System to rest exclusively under FDA control while allowing FDA to partner on those aspects—research on methods and IT infrastructure—that are most logical and efficient to be shared as a national resource. We must explore the potential for the Sentinel System ultimately to support many other regulatory activities critical to a modern health-care system and to interface with related activities, such as comparative effectiveness research (see Chapter 32). To the extent possible, Sentinel must support the health communication tools FDA currently uses while promoting the creation of new capabilities to reach new audiences in innovative ways (see also Chapter 43).

Medical product development and use is a global enterprise. Therefore we must effectively partner with active surveillance initiatives being developed globally. Sharing of epidemiologic and statistical methods is certainly a good place to start. Initiatives such as EU-ADR and IMI-PROTECT (Innovative Medicines Initiative–Pharmacoepidemiology Research on Outcomes of Therapeutics by a European Consortium) are now adding to the accumulated knowledge on methodologic approaches for active surveillance and signal generation. Beyond methods development, we must solve how data might be shared, particularly for rare outcomes where data from every available population is needed to capture enough cases. FDA has been regularly conferring with these international colleagues to build these bridges and begin to share lessons learned.

Finally, we must grapple with the critical question of how the nation will support such activities. We cannot do so until we have clearly delineated the roles all stakeholders will play, including regulated industry.

Sentinel will provide the Agency with important new capabilities. And, if implemented thoughtfully, the initiative will have a major role to play in ensuring that the *ehealth revolution* fulfills its promise of addressing the country's broad public health problems.

Acknowledgement

We would like to acknowledge the contributions to this chapter by our FDA colleagues, Nancy Derr, Mitra Rocca, and Tarek Hammad, MD, PhD. Ms Derr provided expert and energetic technical writing input and kept us on track. Ms Rocca enhanced our discussion of medical informatics-related issues. Dr Hammad provided detailed, thoughtful comments that helped to shape the content of the chapter.

References

1. Public Law 110-85. 110th Congress. US Food and Drug Administration Amendments Act of 2007, September 2007.
2. McAdams M, Staffa J, Dal Pan G. Estimating the extent of reporting to FDA: a case study of statin-associated rhabdomyolysis. *Pharmacoepidemiol Drug Saf* 2008; **17**: 229–39.
3. US Food and Drug Administration. *Sentinel Network Public Meeting*, March 7, 2007. http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm113621.htm (accessed July 2011)
4. US Food and Drug Administration. *Sentinel Network Public Meeting*, March 8, 2007. http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm113626.htm (accessed July 2011).
5. US Food and Drug Administration. *Guidance for Industry, Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment*, March 2005. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126834.pdf (accessed July 2011).
6. US Food and Drug Administration. *Guidance for Industry, E2E Pharmacovigilance Planning*, April 2005. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073107.pdf (accessed July 2011).
7. US Food and Drug Administration. *Drugs@FDA*. http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm (accessed July 2011).
8. Hetherington S, Hughes AR, Mostellar M, Shortino D, Baker KL, Spreen W, *et al*. Genetic variation in HLA-B region and hypersensitivity reactions to abacavir. *Lancet* 2002; **359**: 121–2.
9. Mallal S, Nolan D, Witt C, Masel G, Martin AM, Moore C, *et al*. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse transcriptase inhibitor abacavir. *Lancet* 2002; **359**: 727–32.
10. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, *et al*. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 2008; **358**: 568–79.
11. American Health Information Community. *Personalized Healthcare Work Group Recommendations*, June 3, 2008. http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS_0_10731_848353_0_0_18/Personalized%20Healthcare%20Workgroup%20Recommendations%20Letter.pdf (accessed July 2011).
12. IMS Government Solutions. *Evaluation of Timeliness of Medical Product Uptake for Surveillance in Healthcare Databases*, February 24, 2009. http://www.regulations.gov/#!documentDetail;D=FDA-2009-N-0192-0007;oldLink=false (accessed July 2011).
13. Greene SK, Kulldorff M, Lewis EM, Li R, Yin R, Weintraub ES, *et al*. Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project. *Am J Epidemiol* 2010; **171**: 177–88.
14. Centers for Disease Control and Prevention. *Vaccine Safety Datalink (VSD) Project*. http://www.cdc.gov/vaccinesafety/Activities/VSD.html (accessed July 2011).
15. Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K, *et al*. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 2007; **45** (**10** Suppl 2): S89–95.
16. Brown KS, Kulldorff M, Chan KA, Davis RL, Graham D, Pettus PT, *et al*. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1275–84.
17. Brown J, Lane K, Moore K, Platt R. *Defining and Evaluating Possible Database Models to Implement the FDA Sentinel Initiative*, May 2009. Harvard Pilgrim Health Care Institute. http://www.regulations.gov/#!documentDetail;D=FDA-2009-N-0192-0005;oldLink=false (accessed July 2011).
18. Yih WK, Nordin JD, Kuldorff LE, Lewis E, Lieu TA, Shi P, *et al*. An assessment of the safety of adolescent and adult tetanus-diphtheria-acellular pertussis (Tdap) vaccine, using active surveillance for adverse events in the Vaccine Safety Datalink. *Vaccine* 2009; **27**: 4257–62.
19. Walker AM. Orthogonal predictions: follow-up questions for suggestive data. *Pharmacoepidemiol Drug Saf* 2010; **19**: 529–32.
20. Gould L, Walker A, Ryan P, Schneeweiss S, Santinello N, Sacks S. Using databases for both hypothesis generating and hypothesis confirmation: One database or two? *Pharmacoepidemiol Drug Saf* 2009; **18**: S205.
21. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000; **283**: 8–12.
22. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; **316**: 140–4.

23. Weingart SN, Simchowitz B, Padolsky H, Isaac T, Seger AC, Massagli M, *et al*. An empirical model to estimate the potential impact of medication safety alerts on patient safety, health care utilization, and cost in ambulatory care. *Arch Intern Med* 2009; **169**: 1465–73.
24. Almenoff J, Tonning JM, Gould AL, Szarfman A, Hauben M, Ouellet-Hellstrom R, *et al*. Perspectives on the use of data mining in pharmacovigilance. *Drug Safety* 2005; **28**: 981–1007.
25. Hochberg AM, Hauben M, Pearson RK, O'Hara DJ, Reisinger SJ, Goldsmith DI, *et al*. An evaluation of three signal-detection algorithms using a highly inclusive reference event database. *Drug Safety* 2009; **32**: 509–25.
26. Curtis JR, Cheng H, Delzell E, Fram D, Kilgore M, Saag K, *et al*. Adaptation of Bayesian data mining algorithms to longitudinal claims data: coxib safety as an example. *Med Care* 2008; **46**: 969–75.
27. Observational Medical Outcomes Partnership. http://omop.fnih.org (accessed July 2011).
28. US Department of Health and Humans Services. Health Information Technology: Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology; Interim Final Rule. *Federal Register*, vol. **75**, no. 8, January 13, 2010. http://edocket.access.gpo.gov/2010/pdf/E9-31216.pdf (accessed July 2011).
29. Platt, R. *The PRISM (Post-Licensure Rapid Immunization Safety Monitoring) System for H1N1 Vaccine*. http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/BloodVaccinesandOtherBiologics/VaccinesandRelatedBiologicalProductsAdvisoryCommittee/UCM198079.ppt (accessed July 2011).
30. Indiana Health Information Exchange. http://www.ihie.com (accessed July 2011).
31. South Carolina Health Information Exchange. http://www.schiex.org/index.php (accessed July 2011).
32. 45 Code of Federal Regulations 164.
33. Public Law 107-347. 107th Congress. The Federal Information Security Management Act. December 2002.
34. US Food and Drug Administration. *Adverse Event Reporting System (AERS)*. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm (accessed July 2011).
35. VAERS Vaccine Adverse Event Reporting System. http://vaers.hhs.gov/index (accessed July 2011).
36. US Food and Drug Administration. *Manufacturer and User Facility Device Experience Database (MAUDE)*. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/ReportingAdverseEvents/ucm127891.htm (accessed July 2011).
37. Yih WK, Fireman B, Klein N, Kulldorff M, Lewis E, Lieu T, *et al*. Near real-time post-marketing surveillance: the experience of the Vaccine Safety Datalink. *Pharmacoepidemiol Drug Saf* 2009; **18**: S5–6.
38. Centers for Disease Control and Prevention (CDC); Advisory Committee on Immunization Practices (ACIP). Update: recommendations from the Advisory Committee on Immunization Practices (ACIP) regarding administration of combination MMRV vaccine. *MMWR Morb Mortal Wkly Rep* 2008; **57**: 258–60.
39. National Vaccine Advisory Committee. *Report on 2009 H1N1 Vaccine Safety Risk Assessment*, March 25, 2010. http://www.hhs.gov/nvpo/nvac/reports/vsrawg_report_mar2010.html (accessed July 2011).
40. EU-ADR. http://www.euadr-project.org/ (accessed July 2011).
41. Trifiro G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salamé G, *et al*. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf* 2009; **18**: 1176–84.
42. Coloma P, Schuemie M, Trifiro G, Herings R, Hippisley-Cox J, Gini R, *et al*. Opportunities in diversity: merging healthcare databases in Europe for safety signal detection in pharmacovigilance. *Pharmacoepidemiol Drug Saf* 2009; **18**: S184.

CHAPTER 31
Pharmacoepidemiology and Pharmaceutical Reimbursement Policy

Mitchell Levine[1] *and Jacques LeLorier*[2]
[1] Centre for Evaluation of Medicines, Hamilton, Ontario, Canada
[2] CHUM, Hôpital Hôtel-Dieu de Montréal and Department of Pharmacology, University of Montréal, Montréal, Canada

Introduction

Pharmacoepidemiologic and pharmaceutical reimbursement policy are inextricably intertwined (Figure 31.1). Reimbursement policies inevitably have an impact on whether drugs will be used in a population, and if so under what specific circumstances. This of course influences the effects that a drug can have on a potential population. Reciprocally, pharmacoepidemiologic research can have an equally important impact on the decisions that are made within the realm of reimbursement policy. In this chapter we will focus upon the latter influence, identifying the role and opportunities that pharmacoepidemiologic research can play with regards to the actions of reimbursement policy decision makers.

Although there are a wide variety of different public and private drug benefit programs, an element common to most is that the decision to reimburse a particular drug is usually made after careful consideration of the drug's efficacy (see Chapter 37), safety (see Chapter 1), and pharmacoeconomics (see Chapter 38). Although the data that are reviewed by different drug benefit programs are similar, if not identical, reimbursement decisions may be different amongst programs. Discordant reimbursement decisions are simply a reflection of the application of different value systems, and sometimes political interferences, in the decision making process.

Clinical problems to be addressed by pharmacoepidemiologic research

The "clinical problems" referred to in this chapter are actually issues that involve the medication reimbursement decision process. The data provided to reimbursement decision makers may demonstrate good value based upon a pharmacoeconomic model or a health economic analysis that is embedded within a clinical trial. Whether the model or analysis reflects what will actually happen in the real world, if the drug is reimbursed and subsequently used in clinical practice, is a critical issue. When pharmacoeconomic analyses are conducted using clinical trial data they incorporate many assumptions and practices (see Chapter 38) that are embedded in the clinical trial that do not apply to the real world (see Chapter 36). As a consequence, the results of an economic analysis will be dramatically different depending upon whether it uses clinical trial data instead of clinical practice data. Yet it is the expectation that the pharmacoeconomic analyses used in reimbursement decisions will reflect clinical practice even when they have relied upon clinical trial data. The following are some illustrative examples where the absence of population-based drug data would lead to misleading conclusions.

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

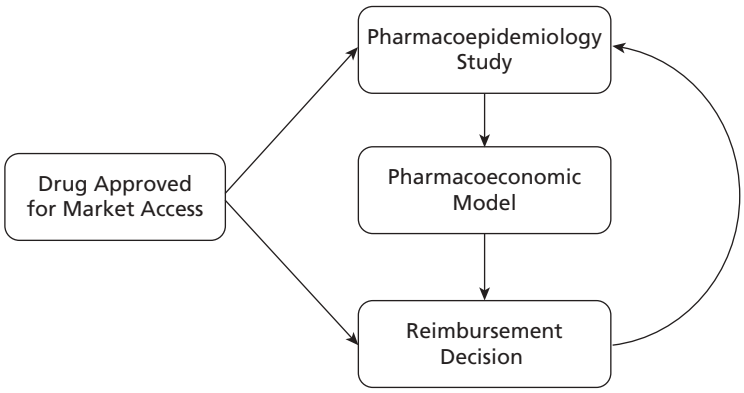


Figure 31.1 Pharmacoepidemiologic studies can provide important data both before and after the initial decision of whether a pharmaceutical product should be reimbursed by a drug benefit program.

Predicting practices and behaviors

In a clinical trial, physician and patient behaviors are often dictated by the research protocol whereas in clinical practice behaviors are far less constrained and are motivated by numerous external influences and personal preferences. For example, it is well recognized that the administration of a non-steroidal anti-inflammatory drug (NSAID) can cause gastrointestinal (GI) ulcer disease in some patients. It is also known that the use of proton pump inhibitors (PPI) can prevent or help heal NSAID-related GI ulcers even while the NSAID is continually administered, and that the co-administration of a PPI can reduce the risk of recurrence in NSAID users.[1] The risk of GI ulcers can also be reduced with the use of selective COX-2 inhibitors as an alternative to the non-selective NSAID.[2] A logical conclusion based on selective COX-2 inhibitor clinical trial data would be to assume that NSAID users who are concurrently using a PPI would be able to discontinue the PPI if they were switched from the non-selective NSAID to the selective COX-2 inhibitor.

In a submission to obtain listing in a drug benefit formulary, an economic model was included that had been extrapolated from clinical trial data and it was proposed that the additional cost of substituting the less expensive NSAID with the more expensive selective COX-2 inhibitor would be offset by the cost reduction associated with the discontinuation of the PPI therapy. But when COX-2 inhibitors were introduced into formularies the use of PPIs did not decrease. Although the exact reasons for the failure to achieve the anticipated cost offset are not known, one consideration is that PPIs were also being used for gastroesophageal reflux disease (GERD), which would not subside when the NSAID was replaced by a selective COX-2 inhibitor.

This example illustrates how randomized clinical trial (RCT) data can fail to predict clinical practices and behaviors that occur when a drug is available in the postmarketing environment. It is only with pharmacoepidemiologic data collected in the world of usual and customary practice that clinical practices and behaviors can be properly assessed, and then incorporated into pharmacoeconomic models for advising reimbursement decision making.

Predicting what the new drug will replace when introduced into the market, and how is it going to be used

When a new drug product is introduced into the market and is reimbursed as part of a drug benefit plan, it will either replace an existing product(s) or become an add-on to current therapeutic practices. This has important implications for the determination of the cost-effectiveness of the new product.

For example, if a new antiplatelet agent that inhibits the adenosine diphosphate (ADP) receptor

subtype $P2Y_{12}$ is approved for use in the prevention of coronary artery thrombosis, the cost-effectiveness (and value) of the drug will be different if: (i) it is used in clinical practice as an alternative to a comparably priced and similarly effective ADP inhibitor, (ii) it replaces a less expensive antiplatelet drug such as acetylsalicylic acid (ASA), or (iii) it is used as a second drug (an add-on) with one of the other drugs. While the regulatory approved indication that is described in the drug product's label may provide some insight into the potential market share of the new drug, the actual use and displacement of other drugs is impossible to predict because there are many variables to consider. In addition, the influence of the marketing efforts by the manufacturer, clinical trial publications, and the clinical experience that practitioners obtain when prescribing a new drug will also influence drug product preferences and subsequent prescribing decisions.

Pharmacoepidemiologic data, by quantifying drug use over time and tracking switching patterns (see Chapter 24), can provide a realistic assessment of market share and the consequences that are relevant to pharmacoeconomic questions and reimbursement decision making. The data can determine if an economic model that has been used to obtain reimbursed market access had used the wrong comparator(s) in the analysis and led to erroneous inferences regarding value and impact.

Cost impact and affordability

Drug reimbursement decisions that are based upon pharmacoeconomic data should ideally be devoid of cost containment and budget impact concerns, as the objective of pharmacoeconomic analyses is to obtain the best value from resources committed towards a program. Nevertheless, sometimes reimbursement decisions are constrained by affordability and whether there is sufficient budget available to support the addition of a specific therapeutic intervention. Under these conditions, it may be necessary to evaluate the financial consequences of including a new drug product on the list of reimbursable items (i.e., the formulary). Pharmacoepidemiologic data pertaining to the use of targeted medications and clinical consequences can be used to evaluate budget impact and affordability of program following a reimbursement decision.

Methodologic problems to be addressed by pharmacoepidemiologic research

The methodologic problems that can be addressed by pharmacoepidemiologic research relate to the internal validity, precision, and the applicability of pharmacoeconomic analyses, the foundation for reimbursement decisions. In a pharmacoeconomic model, the model operates on a number of assumptions that cannot be tested *a priori*. But following a reimbursement decision, those assumptions can be tested empirically using pharmacoepidemiologic data. The opportunity to test the validity of assumptions used in a pharmacoeconomic model has allowed drug programs to offer conditional reimbursement status, whereby pharmacoepidemiologic data is collected after the initial decision, and, if the drug's use does not unfold as expected, reimbursement status can be withdrawn. For instance, when the first PPI, omeprazole, was considered for reimbursement a pharmacoeconomic analysis had predicted, based on clinical trial data, that the duration of treatment for duodenal and gastric ulcers would be 50% of that necessary when an H-2 blocker was used to treat this same conditions.[3] In Quebec, the drug was thus conditionally accepted. However, when a subsequent drug use study was performed by the formulary committee, it demonstrated that in most cases physicians prescribed omeprazole for the same period of time that they had been using for H-2 blockers. Furthermore, omeprazole was mostly used to treat GERD, and the ulcer share of the market was relatively small. It was thus obvious that the savings based on shorter periods of administration predicted by the economic models did not occur in real life. The consequence from the pharmacoepidemiologic study was that omeprazole was transferred to a restricted list which limited its use to ulcers and GERD not responding to the usual doses of H-2 blockers.

Drug use studies are not only useful in correcting formulary decisions, but can also be used to validate current practices. For example, concerns had been expressed that the use of statin therapy in Canada was mostly for primary prevention, an indication of less certain value when compared to the more cost-effective use as secondary prevention of cardiovascular outcomes. A drug use study was conducted and demonstrated that over 50% of statin use was either for secondary prevention or in patients with diabetes. Further, even the majority of primary prevention patients were at the level of moderate risk for a cardiovascular event due to concomitant cardiovascular risk factors.[4] This reaffirmed the acceptable value of statin reimbursement and resulted in no formulary changes or restrictions.

The results of a pharmacoeconomic study are presented as an incremental cost effectiveness ratio (ICER) in units such as cost per QALY (quality adjusted life year) or cost per clinical outcome avoided (see Chapter 38). Because there are two independent variables in the ICER (cost and health outcome) the traditional measure of precision, the 95% confidence interval, does not work particularly well in this circumstance. The "interval" is a two dimensional ellipse that in the vast majority of analyses will include both types of dominant outcomes, that is cost saving with better outcomes and greater cost with worse outcomes. Thus, it will frequently appear that the analysis offers no guidance to the reimbursement decision maker since it cannot distinguish whether the new intervention/therapy is highly desirable or should definitely be avoided. Inserting real clinical (exposure and outcome) data into the model can provide an empirical estimate of the cost effectiveness that simple statistical analyses cannot do. This is because empirical data have no imprecision, that is no confidence interval. The data simply reflect what has happened in the population of interest (rather than an estimate from a sample of the population).

To illustrate, imagine a pharmacoeconomic model that used clinical trial data showing that 2-year survival with a new oncology drug was 25%, with a 95% confidence interval of 17 to 33%. The latter reflects the limits of precision for the results based on the concept that the study subjects are a sample of a much bigger population. Thus the economic model must incorporate the idea that the benefit can be as small as 17% or possibly as large as 33%. This will result in the ICER that is lacking precision. But if a study is conducted amongst *all* the patients in a drug benefit program that have used the drug and the 2-year survival was 23%, that is a number that can be used in the pharmacoeconomic model without using any confidence interval, and as a result the ICER will be more precise.

When pharmacoeconomic models are built, the designated outcomes usually are limited to events known from clinical trial experiences. Because pharmacoepidemiology evaluates drug exposure and outcomes in the postmarketing environment, an iterative approach to economic analysis can be conducted whereby the economic model can be refined and re-analyzed following input of new outcomes (or new rates of known outcomes) that are either beneficial or harmful, which are identified from pharmacoepidemiologic data. For example, the thiazolidinediones (TZDs) were reimbursed on the assumption that improvements in glycated hemoglobin (HbA1c) observed in phase III trials would translate into a decrease in cardiovascular outcomes. It was also assumed that the only adverse effects would be those observed in the Phase III trials. Eventually, additional information from pharmacoepidemiologic studies demonstrated the existence of adverse effects such as cardiovascular events and fractures and that the expected benefits on prevention of cardiovascular events has never been demonstrated.[5,6] This evidence clearly demonstrated that the pharmacoeconomic model was not pertinent to the real-life use of these drugs. It is also possible to test the model using more relevant outcomes than were available from clinical trials. For example, if an antidiabetic agent is evaluated in a clinical trial with the HgbA1c being used as a surrogate outcome for cardiovascular complications associated with diabetes, the pharmacoeconomic model is initially limited to using that surrogate outcome. However, with the conduct of pharmacoepidemiologic studies evaluating cardiovascular endpoints, it may be possible to redesign the model to directly incorporate the real outcome of interest, that is cardiovascular events.

Pharmacoeconomic analyses that rely upon clinical trial data are often restricted to inadequately short durations of exposure and the lack of long-term outcomes. Further, models will often extrapolate from 1- or 2-year data (from a clinical trial) into a model that projects out for 5 or 10 years, or even longer. The validity of employing that type of extrapolation remains to be scientifically demonstrated. Sometimes even within 6 or 12 months the effects of a chronic therapy can change, and unless a study is conducted for a longer period of time the trial results will mislead the pharmacoeconomic analysis. For example, when a pharmacoeconomic model pertaining to the use of etanercept in the treatment of rheumatoid arthritis was conducted by extrapolating from 6 months of clinical trial data the results were much more optimistic from a cost-effectiveness perspective than when the analysis was conducted using pharmacoepidemiologic data collected during a 1-year cohort study.[7]

Pharmacoepidemiology can also be used to demonstrate the unintended and potentially harmful consequences of restricting access to a new drug. In Quebec, clopidogrel was initially restricted to patients who had received a coronary stent. Upon discharge from the hospital the attending physician had to fill in and send to the drug plan a specific form in order to obtain reimbursement for clopidogrel as an outpatient. The problem was that the attending physician who discharged the patient from the hospital was often not an interventional cardiologist who inserted the stent and was not aware of the existence of this form. Because of concerns that this bureaucratic hurdle might result in some patients not continuing on clopidogrel as outpatients, a pharmacoepidemiologic study was done on the RAMQ (Régie de l'assurance maladie du Québec) databases. The assessment showed that an unacceptably large number of patients (20%) were either not continuing on clopidogrel after discharge or were delayed in receiving the drug, putting them at risk for incremental morbidity and mortality.[8] When the drug plan simplified the requirements for clopidogrel reimbursement, the percentage of stented patients who did not receive clopidogrel after discharge decreased to less than 2%.

Another pharmacoeconomic issue that would benefit from pharmacoepidemiologic data involves the international generalizability of economic analyses. When an economic model is constructed for use in one country, applying it to the reimbursement decisions in another country is more complicated than simply a conversion of currency and the determination of the applicable local costs. Practice patterns are different in different jurisdictions and how the resources are used in response to specific events can strongly influence the results of an economic model. For example, in the management of acute coronary syndrome, the threshold for conducting an angiogram or angioplasty procedure can be different among countries. In addition to having a potential impact on clinical outcomes, this would also have a cost impact when making a comparison between two treatments that might have a differential effect on the development of acute coronary syndrome symptoms. Preventing new symptoms would result in greater cost savings in a jurisdiction where the response to new symptoms would have ordinarily generated a greater use of resources. Therefore, in circumstances where international data have not been collected, before applying an pharmacoeconomic model from one country to another it would be desirable to at least conduct a pharmacoepidemiologic study in the new country to determine the nature and scope of applicable clinical practice behaviors and the use of resources.

Currently available solutions

Drug benefit programs in many countries (whether they be private or government-run programs) require pharmacoeconomic analyses to guide reimbursement status of newly available drugs. In these same jurisdictions, large administrative databases exist whereby drug use studies and drug-consequence studies can be conducted in a real world setting (see Part IIIB). To conduct these studies the data must first be anonymized to protect patient confidentiality (see Chapter 35). There must be sufficient computing resources to handle extremely large data sets, and there is a need for skilled human resources to clean up the data as the data would have been originally collected for payment purposes, and not designed for use in research. In the US there are many private insurers

and HMOs with suitable data for conducting analyses (see Chapters 12 and 13). In addition, Medicaid data in the US can also be used to enhance pharmacoeconomic information (see Chapter 14). In Canada, many provinces (which act like large HMOs) have drug claim databases that researchers can access and there is also access to private drug plan data, although the latter cannot be readily linked to clinical outcomes (see Chapter 17). In Europe, many countries have administrative drug data sets that can help guide regulatory authorities that have pharmacoeconomic interest (e.g., NICE, see Chapter 18).

In addition to administrative drug benefit databases there are other means to evaluate drug use and consequences in the postmarketing environment. For example, the Pharmacy Medication Monitoring Program in Canada identifies patients filling a targeted drug (or drug class) in a community pharmacy and conducts standardized telephone interviews with these patients.[9] The program can also send a fax-back questionnaire to their prescribing physician. The data collected provides information regarding how a drug is used in a community-based environment. With the increasing adoption of computerized health records in clinical practice the aggregation of records through primary care networks also offers a method for evaluating the validity and applicability of pharmacoeconomic models (see Chapters 15 and 30).

The future

There are essentially two key questions when making a reimbursement decision: "How is this drug going to be used?" and "What (if anything) is this drug going to replace?" It is not possible to answer these questions on the basis of Phase III trials, especially ones involving a placebo comparator. Furthermore, despite the best efforts at modeling pharmacoeconomic analyses (see Chapter 38), without correct inputs one cannot predict what value will be attained in a postmarketing world. In our view, the best solution for drug reimbursement policy is *conditional* formulary acceptance. This means that the drug would be reimbursed until further population-based studies, which ideally should be financed by the manufacturer of the drug, demonstrate what the drug's use would be like in the real world.

References

1. Lanza FL. A guideline for the treatment and prevention of NSAID-induced ulcers. Members of the Ad Hoc Committee on Practice Parameters of the American College of Gastroenterology. *Am J Gastroenterol* 1998; **93**: 2037–46.
2. Deeks JJ, Smith LA, Bradley MD. Efficacy, tolerability, and upper gastrointestinal safety of celecoxib for treatment of osteoarthritis and rheumatoid arthritis: systematic review of randomised controlled trials. *BMJ* 2002; **325**: 619.
3. Hui WM, Lam SK, Lau WY, Branicki FJ, Lok AS, Ng MM, Lai CL, Poon GP. Omeprazole and ranitidine in duodenal ulcer healing and subsequent relapse: a randomized double-blind study with weekly endoscopic assessment. *J Gastroenterol Hepatol* 1989; **4** (Suppl 2): 35–43.
4. Farahani P, Levine M, Gaebel K, LeLorier J, Perreault S, Gillis J, *et al*. Exploring patient demographic and clinical characteristics associated with lipid-lowering pharmacotherapy use in primary care. *Clin Invest Med* 2007; **30**: E63–E69.
5. Kaul S, Bolger AF, Herrington D, Giugliano RP. Thiazolidinedione drugs and cardiovascular risks. a science advisory from the American Heart Association and American College of Cardiology Foundation. *J Am Coll Cardiol* 2010; **55**: 1885–94.
6. Murphy CE, Rodgers PT. Effects of thiazolidinediones on bone loss and fracture. *Ann Pharmacother* 2007; **41**: 2014–18.
7. Farahani P, Goeree R, Levine M. A comparison between integrating clinical practice setting and randomised controlled trial setting into economic evaluation models of therapeutics. *J Eval Clin Pract* 2006; **12**: 463–70.
8. Sheehy O, LeLorier J, Rinfret S. Restrictive access to clopidogrel and mortality following coronary stent implantation. *Can Med Assoc J* 2008; **178**: 413–20.
9. Willison DJ, Gaebel K, Levine MAH, Borden EK, Poston J, Goldsmith CH, *et al*. Experience in the development of a pharmacy-based postmarketing surveillance network: The pharmacy medication monitoring program. *Ann Pharmacother* 1995; **29**: 1208–13.

CHAPTER 32
Comparative Effectiveness Research

Brian L. Strom, Rita Schinnar, and Sean Hennessy
Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

The history of comparative effectiveness research in the US

The desire to increase the use of scientific evidence to inform clinical decisions has been longstanding. Efforts to make clinical care more rational have been characterized at different times by labels such as *outcomes research*, *effectiveness research*, *evidence-based research*, *health technology assessment,* and, most recently, *comparative effectiveness research* (CER).[1] To combat the perception that the main agenda behind the push for comparative effectiveness research is to drive down the cost of health care, at least one government agency has begun re-branding CER as *patient-centered health research*.[2] As intended by its proponents, the primary purpose of CER is not for making reimbursement decisions (i.e., determining coverage by public and private payers), nor for decisions on which drugs to include or exclude from hospital formularies, nor for influencing practice (clinical) guidelines, nor for influencing other policy initiatives. While results from CER may be useful for any of these areas,[3] the major objective of CER is to provide scientific information to patients and clinicians to assist in health-care decisions. In this chapter, we will continue to use the more commonly accepted term at the time of this book's publication, comparative effectiveness research (CER).

Earlier government initiatives for effectiveness research in the US were attempted first by the Congressional Office of Technology Assessment (established in 1972), then by the National Center for HealthCare 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).[4] The stated objectives of these initiatives were to provide information through studies of patterns of care, to identify optimal treatments, and to achieve economic savings.[5,6]

The next significant political development was the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. MMA authorized AHRQ to support research in the form of systematic reviews and syntheses of the scientific literature, with focus on "outcomes, comparative clinical effectiveness, and appropriateness of health care items such as pharmaceuticals and health care services, including prescription drugs" and including the manner in which they are organized, managed, and delivered.[7] This legislation also provided substantial funding to support this effort.[5] Similar independent initiatives were taking place in the Department of Veterans Affairs (VA) and in the Centers for Medicare and Medicaid Services (CMS), each of which developed internal programs of comparative effectiveness research over this period.[5]

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

Concurrent with these government efforts, the private sector also initiated evidence-based medicine projects (e.g., Blue Cross/Blue Shield Technology Evaluation Center; Emergency Care Research Institute; Hayes, Inc.; and others).[8,9] CER publications have been increasing since at least the early 1980s.[10]

The latest impetus for this effort came from the American Recovery and Reinvestment Act (ARRA Stimulus) of 2009, with an appropriation of $1.1 billion "to study the comparative effectiveness of healthcare treatments".[11] This funding was to be distributed by the AHRQ ($300 million), by the National Institutes of Health (NIH) ($400 million), and by the Office of the Secretary of Health and Human Services ($400 million).

To advance this effort, Congress created the Federal Coordinating Council for Comparative Effectiveness Research to coordinate CER across the Federal government. The 15 members on the Council were selected from a spectrum of federal agencies: Agency for Healthcare Research and Quality (AHRQ); Centers for Disease Control and Prevention (CDC), Office of Strategy and Innovation; Centers for Medicare and Medicaid Services (CMS), Center for Medicare Management; Department of Defense, Office of the Assistant Secretary of Defense for Health Affairs; Food and Drug Administration (FDA); Health Resources and Services Administration (HRSA), HIV/AIDS Bureau; Mental Health Services Administration (SAMHSA), Office of Applied Studies in the Substance Abuse; Office on Disability/ Office of the Secretary at HSS; Office of Health IT Adoption Office of the National Coordinator; Office of the Assistant Secretary for Planning and Evaluation in HHS; Office of Management and Budget; Office of Minority Health; National Heart, Lung, and Blood Institute at the National Institutes of Health; and Veterans Administration.[12] In addition, Congress directed the Institute of Medicine (IOM) to formulate national priorities for CER in order to guide the allocation of research funds productively.[13] On June 30, 2009, both the Federal Coordinating Council and the IOM committee released their reports, providing a working definition of CER and a list of 100 research priorities.[13] Also in June 2009, legislation was proposed (S.1213) to create a Patient Centered Outcomes Research Institute (PCORI) for the purpose of establishing and implementing a national agenda for comparative effectiveness research projects. It was envisioned to be a private, non-profit corporation, funded by public and private sources through a Patient-Centered Outcomes Research Trust Fund. PCORI, discussed more below, is expected to have a budget after its start-up period of over $500 million/year.

History of comparative effectiveness research outside the US

CER has also been adopted formally by the governments of other countries, primarily in Europe.[5] The National Institute for Health and Clinical Excellence (NICE) in the United Kingdom (UK), created in 1999, represents one model for using CER to inform policy research and practice.[14] To help reduce variation in clinical practice and to standardize the quality of care, the primary mandate for NICE has been to evaluate health technology, surgical and diagnostic procedures, and public health interventions for disease prevention, and to develop evidence-based clinical guidelines for these.[14] Other NICE mandates are to identify gaps in knowledge, recommend priorities for research, and help accelerate access to promising new technologies. To improve efficiency, NICE's mandate was recently expanded to include data collection and evaluation of the comparative effectiveness of diagnostics and medical devices, including cost effectiveness.[14] For its evaluations, NICE's advisory committees use objective evidence provided by academic institutions in the UK under contract with NICE to perform evidence syntheses and to conduct small-scale studies entailing primary data collection.[14] NICE has provided a substantial number of evidence-based guidelines for clinical practice,[5,15] though not without controversy and challenge.[16] Despite the criticism, NICE has wide-ranging influence over provider payment levels, development of clinical guidelines for clinical practice, and development of criteria for provider quality assessment and accreditation.[14] 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.[14]

A six-country comparison by Sorenson[17] and a similar three-country comparison by Evans[18] illustrate the considerable efforts extended by European governments to incorporating CER into health policy decisions and the different approaches used for organizing these efforts. In France (the National Authority for Health–Haute Authorité de Santé or HAS, established in 2004[19]), in Germany (the Institute for Quality and Efficiency in Healthcare–Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen or IQWiG, established in 2004[20]), and in the Netherlands (Commissie Farmaceutische Hulp–CHF, which is the 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 to 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.[17,18] Cost-effectiveness data are formally incorporated in evaluations and recommendations about coverage and pricing by most CER entities (UK, Germany, the Netherlands, Sweden).[17] As also characterized by Sorenson, countries differ on the degree to which they "produce" CER, that is conduct evidence synthesis, systematic reviews, and clinical and economic studies (in the UK, Germany, Sweden) or "use" existing CER, relying principally on evidence submitted by the manufacturers (in Denmark, France, the Netherlands).[17] In all countries, the requirement for data on comparators is spelled out and is very specifically defined.[17] Thus, the UK NICE uses current best alternative or routine treatment as comparator. The French HAS requires three comparators from the same therapeutic group: most frequently used; cheapest; and most recently added to the positive list. The Swedish TLV also requires three comparators from the same therapeutic group but different ones: routine treatment, non-medical intervention, and no treatment. The German IQWiG uses either the most effective treatment, or the most widely used, or routine treatment as comparators. The Dutch CHF uses routine treatment as comparator.[17]

Another five-country comparison by Levy *et al.*,[21] which also included Canada and Australia (the Pharmaceutical Benefits Advisory Committee, or PBAS, established in 1987[22]), 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 get included in the formulary despite receiving government funding. Members of these committees are primarily health professionals, with only Canada, Australia, and Scotland including also public representatives, and only Scotland permitting industry representation as well.[21] Other features common to the CER entities in these countries include a process that is: responsive (i.e., transparent, fair, and with reasonable turn-around time); a structure favoring separate entities handling the evaluation of drugs and of medical technologies; and evaluations that ultimately are linked to reimbursement decisions.[21] In all the countries surveyed by these studies,[14,18,21] the CER entities explicitly favor comparative evidence obtained from randomized clinical trials.

Examples of the contributions of some of these CER entities include IQWiG's assessment that insulin analogues (or insulin receptor ligands) such as NPH insulin, lispro insulin, aspart insulin, and glulisine insulin were not superior to human insulin for the treatment of type 1 diabetes;[23] the NICE clinical practice guidelines for post-myocardial infarction prophylaxis[24] and for type 1 diabetes;[25] the HAS guidelines for the management of vitamin K antagonist treatment in different at-risk situations;[26] and various methodologic papers addressed by these CER entities.[27,28]

While the use of CER by government agencies has been well established for a longer period outside the US, much of the recent activity is within the US, and that will be the focus of this chapter.

Definition of effectiveness

As discussed in Chapter 37, a study of *drug efficacy* investigates whether a drug *has the ability* to bring about the intended effect. In other words, 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 *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 ingest it, it will not be effective. Thus, an efficacious drug may lack effectiveness.

Definitions of comparative effectiveness research

The US Congressional Budget Office report of December 2007 defined CER as: "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".[29]

The IOM report later defined comparative effectiveness research as "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".[30]

The report of the US Federal Coordinating Council, described above, proposed a similar definition: "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".[31]

By these definitions, CER includes three key elements: (i) evidence synthesis (identifying and summarizing already extant 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 modifying patient care). Its key elements include the study of effectiveness, rather than efficacy, and that it compares among alternative strategies.

Another key element invoked by the recent US CER initiative is that of inclusiveness of participants in the process. As conceived in the IOM's recommendations for "a robust national CER enterprise,"[30] 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 a process that continuously includes the participation by caregivers, patients, and consumers to provide input regarding issues of public concern.[30] According to Slutsky *et al.*,[32] priorities for CER research must be based on input from all stakeholders in health care; research and synthesis must apply to a wide range of health-care services; and the results must be made accessible to multiple audiences. However, most other countries have excluded the vendors of the technologies being evaluated from the CER process, to avoid commercial bias.

To date, most CER has dealt with the effects of medications, which is one reason why the field is of great interest to pharmacoepidemiologists. However, the full scope of CER is much broader. It addresses the continuum of medical interventions, including drugs, biologics, devices, medical procedures, technologies, prevention strategies, behavioral changes, talk therapies, diagnostics, and health-delivery systems.[30,33] It encompasses beneficial and adverse outcomes as well as economic implications. It focuses attention not only on knowledge creation but also on strategies for implementation (e.g., clinical decision support, see Chapter 25). It urges the use of new data, new analyses of existing data, and systematic reviews of research reports (published or unpublished).[34] It also, as characterized by Lauer,[13] needs to rely on a large number of different research designs; should include very large sample sizes; should utilize an array of technologies that enable quality and efficient health-care delivery; and should account for

the wide range of infrastructures of integrated health-care systems.

Clinical problems to be addressed by pharmacoepidemiologic research

In the context of pharmacoepidemiology, CER completes the path that starts with bench research (characterized by preclinical research to qualify for Phase I regulatory approval), then progresses to bedside research (characterized by proof of concept and efficacy research to qualify for Phase II regulatory approval), and ends with population research (characterized by clinical efficacy to qualify for Phase III regulatory approval) and then with research on the effect of policies (characterized by postmarketing surveillance and pharmacoeconomics research). CER is in one sense narrower than pharmacoepidemiology because it places emphasis on "head-to-head" comparisons of the safety and benefits of treatments and diagnostic strategies, to identify "best-in-class" treatments, in the real world,[5] whereas pharmacoepidemiology often compares users to non-users. On the other hand, CER extends beyond pharmacoepidemiology because it can include non-drug interventions, comparing not only similar treatments but also different diagnostic tests, care delivery systems, etc.

The goals of CER are therefore: (i) to inform decisions among alternative clinical options, (ii) to put new technology into proper perspective versus older technology, (iii) to increase use of more effective clinical options and decrease use of less effective treatments,[1,34–36] and (iv) to identify subgroups of patients more likely to respond to some treatments than others.[37] A consequence of achieving these goals could also be a reduction in health-care costs through avoidance of treatments that do not work or are less effective than alternatives.

The importance of CER is highlighted by current deficiencies in health care. According to a 2007 IOM report,[38] "the rate with which new interventions are introduced into the medical marketplace is currently outpacing the rate at which information is generated on their effectiveness and circumstances of best use,"[4] and "less than half of all medical care is based on or supported by adequate evidence about its effectiveness."[38] In addition, wide variation in practice[39–43] as well as geographic variations in utilization of certain treatments and procedures[44–49] would appear to suggest a lack of "sufficient evidence to determine which approach is most appropriate."[38]

The major gaps in the current knowledge base about treatment interventions, according to Slutsky,[32] are: lack of information about how a treatment works in actual clinical practice, in contrast with how it works in the contrived 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. The disparities in utilization (which could be due to under-utilization, over-utilization, or both) of various treatments, observed among hospitals, providers, and geographic locations, may be attributed partly to these knowledge gaps, which limit the ability of providers and patients to make informed choices among alternative treatment options.[32] This can lead to treatment decisions that are not effective for particular patients in particular circumstances, therefore limiting therapeutic effectiveness and driving up the costs of health care because they are inefficient.

As envisioned in the IOM's report,[30] CER investigations will aim to redress these information gaps, while including all relevant stakeholders and decision makers. However, the inclusion of all relevant stakeholders and decision makers in the process of priority setting, study design, and implementation of results, as advocated by both the IOM[30] and the Federal Coordinating Council for CER,[31] may be over-simplistic and possibly counterproductive. The involvement of participants from such range of backgrounds, orientations, and value judgments may create a cumbersome process, bogged by conflict, and likely to prolong the discovery process and introduce commercial bias. How these conflicting goals, inclusiveness versus expediting unbiased results, will be reconciled remains to be resolved.

As envisioned in the IOM's report,[30] CER investigations will "compare at least two viable

alternative interventions, each with the potential to be 'best practice' ", and will consider how interventions are implemented in usual practice, which includes co-interventions and practice preferences. However, as noted in the report,[30] in the clinical context there may be situations where "watchful waiting" is a reasonable strategy.

As also envisioned in the IOM's report,[30] CER investigations will measure benefits and harms that are important to patients and will produce "results at the population and subgroup levels, and clinical prediction rules to identify patients likely to benefit from an intervention." Traditional efficacy studies typically report average effects, disregarding variability in patient responses. However, providers must make decisions about treatment choices for patients whose profiles are not comparable to the average of study participants in the efficacy trials. An important goal for CER is to explore heterogeneity, looking for subgroups of patients who benefit more (or less) from a given intervention. CER should be used to explore patient variability, and with advances in molecular genetics, this may become increasingly possible.[30] Developments in molecular biology and genomics will increasingly make it possible to determine genetic variation in individual responses to different treatment interventions, with the goal of individualized and predictive medicine[30] (see Chapter 34).

To increase generalizability, CER should include data from patients and physicians from a wide range of care settings. Traditional clinical trials are typically conducted by investigators affiliated with tertiary care hospitals. The vision for CER is that it will provide opportunities for community hospitals and practices to become involved.[50]

Methodologic problems to be addressed by pharmacoepidemiologic research

Issues for evidence synthesis

The synthesis of new and existing information features prominently in definitions of CER. Meta-analysis specifies an approach for performing structured, systematic reviews of published data and for the synthesis of analytic findings by means of formal statistical analysis, to provide an overall measure of effect as a summary of those findings, when appropriate[51,52] (see also Chapter 40). A primary advantage of combining results is the increased precision of estimates and increased statistical power to detect significant effects of treatments, by virtue of the large number of patients pooled across studies. The advantage of large numbers from studies based on multiple population groups also makes possible the detection of effects in subgroups of patients. Meta-analysis also considers variation in study design, study settings, and methods of analysis, which can help explain sources of variation among study findings and some contradictory conclusions in the literature. To be most useful, meta-analyses should include a broad range of data sources, including any valid unpublished data.[53] Also, some feel that systems for evidence rating will need to be developed to cover multiple domains, such as risk of bias, consistency, directness, and precision, dose–response association, presence of confounders that would diminish an observed effect, strength of association, and publication bias.[54]

However, the strengths of meta-analyses are mitigated by their own methodologic problems. Simmonds *et al.*[55] identified several sources of disagreement among experts that can affect the summary findings of meta-analyses: which studies to include or exclude from a meta-analysis, which outcome end-points to consider, and how to pool studies that differ in design and method. In addition, any limitations of the original studies will obviously influence conclusions from the analysis of the pooled studies. Consequently, some have argued that the outputs of meta-analyses may not provide greater insights than the results of individual studies.[55]

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.[55–57] However, issues of access, privacy, and ownership of original data make it difficult for investigators to obtain individual-level data. 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,[58] or because of differences in values and orientations held by different investigators.

The AHRQ and IOM published recommended standards for performing and reporting systematic reviews.[59,60]

The expertise and the effort required to perform a well-conceived and credible meta-analysis are not trivial. Yet, 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.[61]

The value of meta-analyses may also be seriously limited by publication bias. Publication bias can take several forms.[62] Studies with statistically non-significant or negative results are less likely to be published. Studies with statistically significant results and with stronger treatment effects tend to be published with less delay than studies with non-significant results. In addition, some areas of research, such as complementary and alternative medicine, are less likely to be published. Obviously, the summary conclusions from pooled published results tend to be biased because of this preferential selectivity.[62]

A study by Kirtane *et al*.[63] provides an example of a comprehensive meta-analysis that included both randomized clinical trials and observational studies, but analyzed them separately because of the differences in these types of study designs.

Issues for evidence generation

Non-experimental studies

To date, most CER studies have been conducted using non-experimental study designs. Indeed, since CER seeks to study the effects of interventions in the real world, non-experimental approaches can be uniquely useful.[64,65] Further, since one is comparing the effects of different interventions, rather than comparing an intervention to an unexposed group, one is looking to detect smaller differences, and this in turn will require larger sample sizes. Here, too, non-experimental study designs can be logistically more feasible.

Non-randomized study designs, and the methodologic issues they raise,[66] are discussed in detail in much of the rest of this book, and so will be discussed only briefly here. Indeed, most CER studies to date have used the same data resources described in Section III of this book. As noted by the IOM report,[30] 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). Using linked multiple data sources can provide still more powerful tools, enriching the data and enlarging the samples for study. However, the challenge to linking data from multiple sources is the need for standardized code language and format to permit identical computerized queries to be submitted and executed across data resources and standardized format for returning responses from different databases.[67] To be able to access the data and to be able to interpret the analytic findings from the data correctly, familiarity with the logical organization and content of disparate databases is required, including features or quirks in the data that are unique to each database. There are also logistical problems in accessing these data sources, including issues of ownership of data, difficulty in obtaining institutional review board approval, infrastructure, governance, data security, and privacy,[67] and see also Chapter 18.

Non-randomized CER studies of intended effects, however, are more susceptible to confounding by indication than non-randomized studies of unintended effects. As discussed in much more detail in Chapter 37, studies of intended drug effects present a special methodologic problem of confounding by the indication for therapy. In these situations, the risk factor under study is the drug being evaluated and the outcome variable under study is the clinical condition that the drug is supposed to change (cure, ameliorate, or prevent). 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.

Confounding by the indication for the treatment is less of a problem when a study is focusing on unintended drug effects, or side effects, whether they 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 may not be the case for studies of intended beneficial effects, which includes many CER studies. In these studies, one would expect the indication to be more closely related to the outcome variable.

Confounding by the indication for the treatment also may seem less of a problem when doing comparisons among alternative treatments, since both study groups have the indication for treatment. However, this is not to say that non-randomized studies comparing therapeutic alternatives are necessarily free from confounding by indication. The true indication for a treatment is much more subtle than simply the regulator-approved indication. For example, people prescribed a calcium channel blocker as initial treatment for their hypertension are likely to be different from those prescribed a diuretic, with the former being more likely to have pre-existing angina and less likely to have diabetes. Unless choice between the alternatives is effectively random, confounding by indication remains an issue in comparative studies. Indeed, given one is looking to detect differences that are likely to be smaller (compared to studies comparing exposed subjects to unexposed subjects), subtle confounding by indication can be even more problematic.

Considerable effort has been undertaken to develop more effective methods for controlling confounding, for example propensity scores, instrumental variables, etc., in studies based on administrative or other observational data[68] (see also Chapters 37 and 47). However, it is important to keep in mind that most approaches (including propensity scores) are still dependent on identifying and measuring those variables that are the true predictors of therapeutic choice. Identifying and measuring, and thereby adjusting for, everything important is often not possible.[69] Approaches like instrumental variables are promising alternatives. However, finding valid instruments in pharmacoepidemiology is extremely difficult—some would say impossible.[70] Much more work is needed in these areas, to advance the field of CER.

Standards for performing and reporting observational studies were provided by several professional associations.[71–73]

Experimental studies

For the reasons described above, the use of clinical trial designs will always be of paramount importance for generating new evidence on CER. However, CER will focus on using clinical trial designs that are flexible, adaptive, pragmatic, practical, efficient (all the different terms used to refer to this new design), in contrast to the traditional randomized, blinded, placebo-controlled clinical trials[74–77] (see also Chapter 36 for a discussion of large simple trials).

Traditional clinical trials are typically inadequate to address comparative effectiveness.[78] The major limitation of traditional randomized clinical trials is their rigid design specifications and protocols. These include eligibility criteria that exclude patients with co-morbid conditions, pregnant women, or ethnic minorities; specification of only a few study outcomes, and mostly short-term outcomes; and artificial study settings that do not resemble clinical practice in the "real world" and do not resemble patient adherence with recommended therapy regimen in the "real world". Traditional clinical trials are also complex and time-consuming, and thus not feasible for comparing interventions as evidence accumulates over longer periods of time about the natural history of the medical conditions and the treatments studied.[78] In contrast, pragmatic clinical trials include patients with co-morbid conditions and from diverse demographic backgrounds, providers from community settings instead of only tertiary settings, comparator treatments that are in use in clinical practice (rather than placebo controls), variations in the treatment as patients respond differently, and outcomes that matter to patients and clinicians rather than investigators and drug companies.[74,77,78]

These aspects of pragmatic clinical trials represent normal real world conditions. A recent example of conducting a pragmatic trial for CER is the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC) Study (see also Chapter 7), which compared ziprasidone and olanzapine for their risk of non-suicide mortality.[79,80] This randomized trial, discussed more below, examined approximately 18 000 patients in 18 countries, and cost $85 million.

Pragmatic clinical trials raise their own problems. The liberal inclusion criteria characteristic of this study design assures greater representativeness of the study groups, but this increased heterogeneity can decrease the probability of detecting a given treatment effect as statistically significant.[74] The larger samples used by pragmatic trials make them more expensive than smaller trials.[81] Pragmatic clinical trials emphasize evaluation of long-term outcomes, which requires greater resources because of the need to follow up study groups over long periods. Loss to follow-up over time and/or nonadherence over time can introduce bias.[75] The lack of blinding in pragmatic trials creates potential for biased observations and a threat to internal validity.[74] Another limitation of this type of trial results from the flexible treatment protocols preferred. Specifically, these trials involve the participation of community providers in their usual practice. Accordingly, they can vary the treatment process to different patients depending on the variable responses to therapy and they can vary the dose and regimen in the same patients over time. This flexibility makes possible the assessment of the outcomes of the composite treatment, but does not permit the assessment of particular components within the treatment process.[75]

Other limitations are inherent in all clinical trials.[82] Some interventions cannot be investigated with clinical trials because of ethical considerations, even though such trials may be preferred scientifically. Further, the strength of a clinical trial is also one of its weaknesses; it can answer only a very focused question, and there are inevitably many others that also need to be answered, to place a treatment in its appropriate context in our therapeutic armamentarium. Finally, and of great importance, given the cost of clinical trials, it is not practical to undertake them for most CER questions. Thus, much of CER will continue to use the techniques of non-experimental pharmacoepidemiology for the foreseeable future, but their limitations must always be kept in mind in interpreting the emerging results.

Issues for evidence dissemination

Dissemination has several distinct goals. One goal involves identifying priority topics and comprehensively identifying available information on these topics and developing objective interpretations of the information,[83] as provided by the Cochrane Collaboration reviews, for example.[84] The output from this research becomes the source information for dissemination to clinicians, patients, and policy makers. Another goal involves knowledge translation, namely, using this information as the basis for drafting clinical guidelines. The IOM recently proposed standards for developing trustworthy clinical practice guidelines.[85] A third goal involves knowledge exchange and utilization, achieved by the actual distribution of the information and educating clinicians, patients, and policy makers about current knowledge and best practice. A final goal involves monitoring and assessment of whether the above efforts translate into actual good practice and, if not, to identify which means of dissemination have a greater chance to create an impact. Each of these goals presents challenges that will require creative approaches to accomplish successfully.

The first goal will be achieved through expanding efforts on systematic reviews and studies using novel research designs (see above), targeting in particular the priority research areas that were identified by the IOM as having gaps in knowledge. Important, however, will be the creation of a national comparative effectiveness research inventory that is augmented periodically to include newly published studies.[2] To be useful, this inventory should be more than just a depository, but should be constructed to enable searches to identify studies and summaries by numerous dimensions, and it should be easily accessible and usable by interested stakeholders.[2]

The second 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 practical guidelines for clinical practice. To be useful, the guidelines need to be unambiguous. Yet, recommendations by experts may be ambiguous because of disagreement in professional judgment. Ideally, the guidelines also need to be both comprehensive for general patient care and specific for particular patient circumstances—a very demanding specification. Furthermore, to remain relevant, the guidelines need to be updated periodically to incorporate new information about existing interventions and information about new treatments.

The third goal of helping providers implement the clinical guidelines in practice may be attained by more intensive use of technology. Computerized physician order entry (CPOE) systems, supplemented by computerized clinical decision-support systems 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 prescribing co-interacting drugs, etc.) are examples of such tools. Other strategies for achieving knowledge exchange and knowledge utilization will require educating clinicians and patients about what treatments work best.[2,31,86] These efforts should include monitoring to assure that information is integrated into the normal workflow and decision processes of clinicians and patients.

However, the availability of evidence-based knowledge and even availability of published guidelines by professional associations may still not assure translation into clinical practice.[40–49] Therefore, the fourth goal to perform evaluation studies to determine the best strategies for getting through to the practicing providers is also important.[87–89] This will require significant funding. To make certain that providers use the new evidence, AHRQ has planned to spend $29.5 million beginning in Spring 2010 on projects aimed at implementing innovative approaches to integrating CER findings into clinical practice and health-care decision making.[90]

Currently available solutions

Organizational approaches

AHRQ was one of the arms of the US Federal Government selected by the US Congress to distribute a portion of the stimulus funds ($300 million) allocated for conducting CER, and in fact much of the $400 million assigned to the Secretary of Health and Human Services was distributed through AHRQ as well. This mission is carried out by AHRQ's Effective Health Care (EHC) program, established in 2005 to provide understandable and actionable information for patients, clinicians, and policy makers.[91] Historically, AHRQ has used its funding to support non-experimental studies, but with the augmented stimulus funding, some clinical trials have been funded as well.

AHRQ's EHC program has supported, since 1997, Evidence-based Practice Centers (EPCs), which focus on evidence synthesis. These Centers are located at 15 medical schools, universities, or medical centers and encompass medical researchers with diverse scientific backgrounds. These Centers produce comparative effectiveness reviews or effectiveness reviews on medications, devices, and other health-care services, and subject these reviews to an editorial process to ensure the accuracy, quality, consistency, and credibility of their reports.[92] Accordingly, a group of stakeholders and scientists is charged with selecting and prioritizing topics for comparative effectiveness systematic reviews. The group uses seven criteria adopted by the EHC program to identify which topics are important. If the group then determines that a new systematic review of CER on such topics will not duplicate existing research syntheses, and determines that there is adequate research to justify a new review, then it will recommend that new meta-analyses be undertaken.[92] Another characteristic is that this process is designed to be transparent and accountable.[92]

AHRQ's EHC program also has supported the DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) Network, a collection of research centers created in 2005. These centers gather new knowledge and information on specific

treatments. The DEcIDE Network conducts studies on the outcomes, effectiveness, safety, and usefulness of medical treatments and services. DEcIDE uses an unusual funding mechanism of task order contracts, so that project topics can be chosen by AHRQ, applications can be reviewed quickly, and studies conducted in a timely fashion, under a contract mechanism.

AHRQ's John M. Eisenberg Center for Clinical Decisions and Communications Science translates comparative effectiveness reviews and research reports created by the EHC program into short, easy-to-read guides and tools that can be used by consumers, clinicians, and policymakers.

AHRQ's Centers for Education and Research on Therapeutics (CERTs) is a national initiative to increase awareness of the benefits and harms of new, existing, or combined uses of therapeutics (drugs, medical devices, and biological products) through education and research. The CERTs program was a network of 14 research centers, each focusing on a broad therapeutic theme. The program is funded and run as a cooperative agreement by AHRQ, in consultation with the US Food and Drug Administration (FDA). Unfortunately, it has just been reduced to 6 centers.

The NIH has also been very involved in CER. Historically, NIH has used its CER funding to fund mostly clinical trials, but some non-experimental studies were funded as well. Given the scale of NIH (budget over $30 billion/year) and its decentralized organization, there is no way to succinctly summarize all of its work in CER, as most of the 27 NIH institutes/ centers have been involved. One new locus of CER in NIH is the network of Clinical and Translational Science Awards (CTSAs), which have increasingly been interested in such work. The CTSAs prepared a white paper with recommendations for advancing the CER concepts into practice, in particular promoting the translation of results of clinical and translational research into practice and public policy.[93] NIH also funds multiple different clinical trial research networks, for example in HIV/AIDS, asthma, leukemia, drug-induced liver injury, maternal–fetal medicine, and many, many more. As noted above, of the $1.1 billion made available in the Stimulus Act for CER, $400 million was targeted to NIH, supplementing the CER activities it was already doing.

In addition, the new Patient-Centered Outcomes Research Institute (PCORI), created under the Patient Protection and Affordable Care Act of 2010 (Section 6301 and Section 10602, Public Law 111–148)[94] is a non-profit organization designed to serve a role in CER. Specifically, it is charged to fund the conduct of research projects that provide evidence on how diseases, disorders, and other health conditions can effectively and appropriately be prevented, diagnosed, treated, monitored, and managed, with the goal to assist patients, clinicians, purchasers, and policy-makers in making informed health decisions. PCORI members, appointed by the US Comptroller General, have been selected from academia, hospitals, health care industry, and patient organizations, with only two government officials (from NIH and AHRQ) included.[95] The creation of a Methodology Committee of PCORI was also mandated by Congress, to develop standards and methods for the conduct of research, and the translation of research results in a useful and understandable way.[96] PCORI will be funded by the Patient-Centered Outcomes Research Trust Fund, which was allocated $10 million for 2010, $50 million for 2011, and $150 million for 2012. In future years, the Trust Fund will include an annual $2 fee per Medicare beneficiary transferred from the Medicare Trust Fund, and an annual $2 fee per-covered-life assessed on private health plans, adjusted for health expenditure inflation. In total, annual funding for PCORI could be more than $650 million.

Many other organizations are also beginning to participate in this promising new surge of research.[9]

Finally, the question of who should be funding CER studies has given rise to a thorny debate, partly motivated by political ideology and partly by business considerations. Some argue that CER should not be funded by government. Since the goal is to compare the benefits and harms of alternative products or medical technologies, it is argued that sponsorship for such studies belongs in the private sector. The counter argument can be made, however, that if the government will not be funding

such research, this type of research is not likely to take place, or will not be unbiased. In general, as we have seen until now, it is often not in the interest of a manufacturer to risk subjecting its product to head-to-head comparison against a competitor's product, with the possibility of losing out in this competition. Further, such comparisons are less likely to be commercially biased if they are funded, designed, and conducted by organizations without an inherent conflict of interest.

Current applications of CER

Hochman and McCormick[97] conducted a survey of recently published studies of comparative effectiveness, targeted at studies that evaluated existing (rather than novel) medications; compared active therapies (rather than placebo comparators); compared medications with non-pharmacologic 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 non-pharmacologic therapies, emphasizing the need for expanding CER. As noted also by the IOM committee, the US "lacks a national infrastructure for learning from the delivery of health care" through research.[2,30]

Shepherd *et al*.[98] analyzed the comparative effectiveness of different inhaled corticosteroids (used alone and used in combination with long acting beta-2 agonists) for the treatment of chronic asthma in adults and children. This paper represents an example of a systematic review of a large number of randomized controlled trials and systematic reviews to assess the clinical and cost effectiveness of these drugs. It provides a detailed account of the search method used, the data extraction method, the appraisal strategy, the data synthesis approach used, that is both a narrative synthesis and meta-analysis, and the methods used for the economic evaluations. It compared inhaled corticosteroids with each other, as well as combination inhalers with each other and with inhaled corticosteroids alone. It provided extensive summaries of the results, and the limitations associated with this type of review.

In contrast, the ziprasidone observational study of cardiac outcomes (ZODIAC) study by Strom *et al*.[79,80] is an example of real-world research using a randomized, large simple trial of patients with schizophrenia receiving routine medical care in naturalistic practice. Accordingly, the only protocol-mandated intervention in this study was the requirement for random assignment of the participants to one of the study treatment groups (ziprasidone or olanzapine). Beyond this, physicians and patients were free to change regimens and dosing based on patients' responses to the assigned medication; free to use concomitant medications, including other antipsychotics; data collection during the trial was limited so as not to burden the physicians and patients (e.g., changes in dosing were not recorded and electrocardiograms were not required at baseline or during the course of the study); and physicians and patients were not blinded to the treatment allocation.[79]

The future

Funding

The commitment by the US federal government to continue to support comparative effectiveness research was manifest in the President's budget proposal for fiscal year (FY) 2011, with $286 million for Patient-Centered Health Research (the re-branded term for CER) through AHRQ. In contrast, in FY 2010 it had been only $21 million. Of course, the actual funding will depend the actions of the newly installed Congress. As noted above, NIH continues to invest substantially in CER. Substantial new funding will come from the new non-governmental, non-profit Patient-Centered Outcomes Research Institute and its Patient-Centered Outcomes Research Trust Fund, with a budget plan of $50 million for FY 2011, $150 million for FY 2012, then increasing to perhaps over $600 million/year.

Currently, the predominant source of funding for CER in the US is the federal government, while much of the funding for clinical efficacy research

comes from industry sources. In the future, the CER enterprise could be expanded even further if industry were to routinely conduct and sponsor comparative trials with products being proposed for marketing. In this new CER environment, given the results of CER studies will be used by payors and providers, it is possible that companies will also invest resources, allowing them to compete on the utility of their products rather than competing just on their marketing ability.[3]

In view of these prospects for expanded funding of CER, the future of CER seems promising. However, the effectiveness of this program will depend also on building an infrastructure to sustain CER in the long term.

Governance

The Federal Coordinating Council for Comparative Effectiveness Research and the Institute of Medicine developed a strategic framework and priorities for CER. The question for the future is who will control the priorities and allocation of funding for CER going forward. Given the high stakes involved, since the outputs of CER may favor or disfavor available treatments, industry understandably wants to have direct input into the governance process. Industry is especially suspicious of government agencies controlling CER because of their susceptibility to political pressure.

Governance of CER in the US is currently distributed across several federal agencies, such as AHRQ, NIH, and HHS. It has been observed by some that the CER enterprise has been impaired by a lack of coordination, and that the predominance of federal control is undesirable.[5,35] Accordingly, alternative governance arrangements have been proposed. An ideal organizational structure should perhaps be a federally supported non-profit corporation that is somehow sheltered from political influence, yet accountable to Congress.[5,99] Clearly, this is difficult to achieve. However, the recently established PCORI seems to fit this model. Its 21-member board includes representatives from academia, hospitals, patient organizations, state health agencies, industry, and the Directors of the AHRQ and NIH (or their designees). PCORI, in turn, can provide its funding to AHRQ and NIH, or to others.

Regardless of the governance arrangements, the process should be transparent to all stakeholders so that policy, prioritization, and funding decisions are fully credible.

Human capital development

As noted by the Federal Coordinating Council,[31] training will be required of new researchers to apply the specialized methods of CER and to develop CER methods.[100] Specialized skills are needed to perform traditional randomized clinical trials and novel pragmatic trials. Specialized expertise is also needed to perform formal meta-analyses and to perform non-experimental studies, using either *ad hoc* data collection or existing databases. Special training is needed to successfully access and link various databases. Finally, the field needs individuals able to translate the findings into practice guidelines and for dissemination. The CER emphasis on community participation and inclusion will dictate that experts from many different fields and backgrounds will be required to communicate with each other, finding 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 practicing clinicians from specialties relevant to given studies, and including pharmacoepidemiologists.[5] These teams will need to have the capacity to develop a shared understanding of basic scientific terminology and methods.

Thus, it is necessary to develop and support training programs for researchers seeking careers in CER.[101] In addition to preparing a cadre of researchers with expertise in CER methods, there needs to be a critical mass of such researchers to undertake the large number of studies needed to fill current gaps in knowledge and to continuously update the knowledge inventory with systematic reviews. AHRQ and NIH, through their stimulus funding, began to address this, but much more is needed.

Cost containment

A major issue for the future is the use of CER results to control health-care costs. The relevance of including cost-effectiveness analyses in CER

investigations is unquestionable. The natural inclination of third party payers is to use the best evidence available to select cost-efficient interventions and health services for determining reimbursement. Critics of CER, however, fear that this is just what will happen. The results of CER will inevitably affect reimbursement decisions by public and private payers, being used for rationing health services.[3,36,102] However, in fact, health-care rationing already occurs, either implicitly or explicitly; therefore it is better if rationing is based on clinical data.[3]

Regardless, the primary goal of CER is to lead to better care, not necessarily cheaper care. The goal of CER is to find out what works better.[3] As the emphasis is on comparative research, CER is not going to dictate that one not intervene in certain settings, but rather how best to intervene in those settings.[3] In the face of calls for greater accountability by health-care providers, the findings from CER will increase pressure on hospitals to make sure that patients get the most effective care. This could result in abandoning expensive technologies that are no better than less expensive options. However, it also could result in paying for a more expensive technology because the evidence shows it is better.[3,35,101]

Toward that end, it is a fallacy to think that the results of CER can only lead to reduced health-care costs. Since research findings may actually show a more expensive treatment option to be better, CER may lead to increased absolute costs (but, in fact, more cost-effective therapy, see Chapter 38). This is of concern to industry because, when comparing technologies, there will be winners and losers,[3,93] and of concern to third party payors and employers who may be pressured to cover treatments found to be effective by CER even though more expensive.[93]

A complete understanding of the net benefits of one intervention compared with another requires consideration of the relative costs and harms as well. It may be difficult from a national public policy perspective to ignore CER in decisions of which interventions to reimburse. Yet, as noted by Kerridge,[82] decisions based on what is best at the population level may sometimes conflict with decisions made on what is best at the individual level. Balancing the ethical, financial, and scientific efficacy and effectiveness concerns will not be resolved easily or quickly (see also Chapter 38).

The strong response from some CER advocates is that CER should not be used for cost containment decisions,[5,35,103] and that the experts conducting CER research should not be placed in a role of using their findings about treatment effectiveness to recommend reimbursement guidelines. In any event, reimbursement decisions need not be under the authority of the same agencies that support CER.[5] However, undoubtedly, well-done CER will and should inevitably affect reimbursement decisions.

Ultimately, CER is not by itself going to solve the world's health-care spending problem ($2.5 trillion in 2009 and projected $4.4 trillion in 2018, roughly 20.3% of global GNP). In some cases, the studies will find that the more expensive treatment is best. However, over time, it should save money by preventing wasteful spending on treatments that are less effective.[3]

Reasonable expectations of CER

Though the research community is energized, expectations must be tempered by several limits to what CER can realistically solve. It is infeasible to expect that CER can address all therapeutic questions; health care is simply too complex. The practice of medicine is as much the application of art as it is the application of science. If it were possible to base the practice of medicine entirely on science, it would take into account not only complex pathophysiology, but also behavioral factors such as values, perceptions, and attitudes about risks, quality of life preferences, cost tradeoffs, etc. However, as the science underlying current medical practice is 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 science. As stated by Kerridge *et al.*,[82] "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."

Finally, over-emphasis on scientific evidence can lead to therapeutic nihilism, that is a paralysis when such evidence is unavailable. In the face of uncertainty, variation among reasonable but unproven options should be tolerated and even encouraged, as it will facilitate later evaluation. This is 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 underlying future CER studies.

References

1. Luce BR, Drummond M, Jönsson B, Neumann PJ, Schwartz JS, Siebert U. 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. Gaige M. Comparative effectiveness research: A valuable tool for health-care reform? *Penn Medicine* Winter 2009/2010: p. 29.
4. Congressional Budget Office. *Research on the Comparative Effectiveness of Medical Treatments*, 2007. http://www.cbo.gov/ftpdocs/88xx/doc8891/12-18-ComparativeEffectiveness.pdf (accessed Jan 2011).
5. Manchikanti L, Falco FJ, Boswell MV, Hirsch JA. Facts, fallacies, and politics of comparative effectiveness research: Part I. Basic considerations. *Pain Physician* 2010; **13**: E23–54.
6. Roper WL, Winkenwerder W, Hackbarth GM, Krakauer H. Effectiveness in health care. An initiative to evaluate and improve medical practice. *N Engl J Med* 1988; **319**: 1197–202.
7. H.R. 1 The Medicare Prescription Drug, Improvement, and Modernization Act of 2003. Public Law 108–73, enacted December 8, 2003.
8. National Pharmaceutical Council. *A Brief History of Comparative Effectiveness Research And Evidence-Based Medicine*. http://www.npcnow.org/Public/Issues/i_cer/cer_toolkit/A_Brief_History_of_Comparative_Effectiveness_Research_And_Evidence-Based_Medicine.aspx (accessed March 2011).
9. Berry E, amednews.com. *Comparative Effectiveness Research: New ways to say "no"*, Sept 21, 2009. http://www.ama-assn.org/amednews/2009/09/21/bisa0921.htm (accessed July 2011).
10. Honig PK. Comparative effectiveness: The fourth hurdle in drug development and a role for clinical pharmacology. *Clin Pharmacol Ther* 2011: **89**: 151–4.
11. H.R. 1 The American Recovery and Reinvestment Act of 2009. Public Law 111-5, enacted February 17, 2009.
12. *Federal Coordinating Council for Comparative Effectiveness Research Membership*. http://www.hhs.gov/recovery/programs/os/cerbios.html (accessed March 2011).
13. 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.
14. 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.
15. Buxton M. Implications of the appraisal function of the National Institute for Clinical Excellence (NICE). *Value Health* 2001; **4**: 212–6.
16. Poole-Wilson PA. NICE: guessing is not guidance. National Institute for Clinical Excellence. *Heart* 2002; **87**: 510–2.
17. Sorenson C. Use of comparative effectiveness research in drug coverage and pricing decisions: a six-country comparison. *Issue Brief* (Commonwealth Fund) 2010; **91**: 1–14.
18. Evans H. *Comparative Effectiveness in Health Care Reform: Lessons from Abroad*, Feb 4, 2009. http://www.heritage.org/Research/Reports/2009/02/Comparative-Effectiveness-in-Health-Care-Reform-Lessons-from-Abroad (accessed July 2011).
19. Haute Autorité de Santé (HAS) (French National Authority for Health). http://www.has-sante.fr/portail/jcms/c_5443/english (accessed July 2011).
20. Nasser M, Sawicki P. *Institute for Quality and Efficiency in Health Care: Germany*, July 2009. http://www.commonwealthfund.org/~/media/Files/Publications/Issue%20Brief/2009/Jul/Chalkidou/1294_Nasser_CER_Germany_issue_brief_724.pdf (accessed July 2011).

21. 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.
22. Sweeny K. *Key Aspects of the Australian Pharmaceutical Benefits Scheme*, Working Paper No. 35, Pharmaceutical Industry Project, November 2007. http://www.cfses.com/documents/pharma/35-Key_aspects_of_PBS_Sweeny.pdf (accessed July 2011).
23. IQWiG Reports. *Long-acting Insulin Analogues in the Treatment of Diabetes Mellitus Type 1*. Executive summary of final report A05-01 Version 1.0; Cologne, Germany: Institute for Quality and Efficiency in Health Care, 2010.
24. Fox KF; UK National Institute for Clinical Excellence. A critical evaluation of the NICE guidelines for post-myocardial infarction prophylaxis. National Institute for Clinical Excellence. *Expert Opin Pharmacother* 2001; **2**: 2079–84.
25. NICE. *CG15. Diagnosis and Management of Type 1 Diabetes in Children and Young People*. National Institute for Clinical Excellence, July 2004. www.nice.org.uk (accessed July 2011).
26. Pernod G, Godiér A, Gozalo C, Tremey B, Sié P; French National Authority for Health. French clinical practice guidelines on the management of patients on vitamin K antagonists in at-risk situations (overdose, risk of bleeding, and active bleeding). *Thromb Res* 2010; **126**: e167–74.
27. Skipka G, Bender R. Intervention effects in the case of heterogeneity between three subgroups. Assessment within the framework of systematic reviews. *Methods Inf Med* 2010; **49**: 613–7.
28. Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, *et al*. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008; **61**: 857–65.
29. 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 July 2011).
30. Institute of Medicine. *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC: National Academies Press, 2009. http://www.iom.edu/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx (accessed Dec 2010).
31. Federal Coordinating Council for Comparative Effectiveness Research. *Report to the President and the Congress*. Washington, DC: Department of Health and Human Services, June 2009. http://www.hhs.gov/recovery/programs/cer/cerannualrpt.pdf (accessed Dec 2010).
32. Slutsky JR, Clancy CM. Patient-centered comparative effectiveness research. *Arch Intern Med* 2010; **170**: 403–4.
33. Volpp KG, Das A. Comparative effectiveness–thinking beyond medication A versus medication B. *N Engl J Med* 2009; **361**: 331–3.
34. Sox HC, Helfand M, Grimshaw J, Dickersin K; PLoS Medicine Editors, Tovey D, Knottnerus JA, Tugwell P. Comparative effectiveness research: challenges for medical journals. *Trials* 2010; **11**: 45.
35. 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.
36. Martin DF, Maguire MG, Fine SL. Identifying and eliminating the roadblocks to comparative-effectiveness research. *N Engl J Med* 2010; **363**: 105–7.
37. Benner JS, Morrison MR, Karnes EK, Kocot SL, McClellan M. An evaluation of recent federal spending on comparative effectiveness research: priorities, gaps, and next steps. *Health Aff* (Millwood). 2010; **29**: 1768–76.
38. Institute of Medicine. Learning What Works Best: The Nation's Need for Evidence on Comparative Effectiveness in Health Care, Sept 2007. http://www.iom.edu/~/media/Files/Activity%20Files/Quality/VSRT/ComparativeEffectivenessWhitePaperESF.pdf (accessed July 2011).
39. Institute of Medicine. *Knowing What Works in Health Care: A Roadmap for the Nation*, Jan 2008. http://www.iom.edu/Reports/2008/Knowing-What-Works-in-Health-Care-A-Roadmap-for-the-Nation.aspx (accessed July 2011).
40. Tang A, Hooper T, Hasan R. A regional survey of chest drains: evidence-based practice? *Postgrad Med J* 1999; **75**: 471–4.
41. Likosky DS, FitzGerald DC, Groom RC, Jones DK, Baker RA, Shann KG, *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.
42. Goldberger JJ, Bonow RO, Cuffe M, Dyer A, Rosenberg Y, O'Rourke R, *et al*.; PACE-MI Investigators. beta-Blocker use following myocardial infarction: low prevalence of evidence-based dosing. *Am Heart J* 2010; **160**: 435–42.

43. Phillips KA, Marshall DA, Haas JS, Elkin EB, Liang SY, Hassett MJ, *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.
44. Piccini JP, Hernandez AF, Dai D, Thomas KL, Lewis WR, Yancy CW, *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.
45. Matlock DD, Peterson PN, Heidenreich PA, Lucas FL, Malenka DJ, Wang Y, *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* 2010; **4**: 114–21.
46. Sung JC, Curtis LH, Schulman KA, Albala DM. Geographic variations in the use of medical and surgical therapies for benign prostatic hyperplasia. *J Urol* 2006; **175**: 1023–7.
47. Warren OJ, Wallace S, de Wit KL, Vincent C, Darzi AW, Athanasiou T. Variations in the application of various perfusion technologies in Great Britain and Ireland—a national survey. *Artif Organs* 2010; **34**: 200–5.
48. Lougheed MD, Garvey N, Chapman KR, Cicutto L, Dales R, Day AG, *et al*. Variations and gaps in management of acute asthma in Ontario emergency departments. *Chest* 2009; **135**: 724–36.
49. Luo X, Pietrobon R, Curtis LH, Hey LA. Prescription of nonsteroidal anti-inflammatory drugs and muscle relaxants for back pain in the United States. *Spine* 2004; **29**: E531–7.
50. Aston G. Comparative effectiveness (why does it matter to you?) *Hospitals & Health Networks Magazine* 200; **83**: 22–5.
51. Ottenbacher KJ, Petersen P. Quantitative reviewing of medical literature. An approach to synthesizing research results in clinical pediatrics. *Clin Pediatr* 1984; **23**: 423–7.
52. Delahaye F, Landrivon G, Ecochard R, Colin C. Meta-analysis. *Health Policy* 1991; **19**: 185–96.
53. Chou R, Aronson N, Atkins D, Ismaila AS, Santaguida P, Smith DH, *et al*. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. *J Clin Epidemiol* 2010; **63**: 502–12.
54. Owens DK, Lohr KN, Atkins D, Treadwell JR, Reston JT, Bass EB, *et al*. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions—agency for healthcare research and quality and the effective health-care program. *J Clin Epidemiol* 2010; **63**: 513–23.
55. Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials* 2005; **2**: 209–17.
56. Broeze KA, Opmeer BC, Bachmann LM, Broekmans FJ, Bossuyt PM, Coppus SF, *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.
57. Koopman L, van der Heijden GJ, Hoes AW, Grobbee DE, Rovers MM. Empirical comparison of subgroup effects in conventional and individual patient data meta-analyses. *Int J Technol Assess Health Care* 2008; **24**: 358–61.
58. Manchikanti L, Datta S, Smith HS, Hirsch JA. 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.
59. AHRQ. Effective Health Care Program. *Comparative Effectiveness Review Methods: Clinical Heterogeneity*, publication no. 10-EHC-070-EF, Sept2010. http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=533 (accessed July 2011).
60. IOM (Institute of Medicine). *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: The National Academies Press, 2011.
61. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007; **147**: 224–33.
62. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, *et al*. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010; **14**: iii, ix–xi, 1–193.
63. Kirtane AJ, Gupta A, Lyengar S, Moses JW, Leon MB, Applegate R, *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.
64. Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox P, Gliklich R. Why observational studies should be among the tools used in comparative effectiveness research. *Health Aff (Millwood)* 2010; **29**: 1818–25.

65. Teutsch SM, Berger ML, Weinstein MC. Comparative effectiveness: asking the right questions, choosing the right method. *Health Aff (Millwood)* 2005; **24**: 128–32.
66. Strom BL. Methodologic challenges to studying patient safety and comparative effectiveness. *Med Care* 2007; **45** (10 Suppl 2): S13–5.
67. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. *Health Serv Res* 2010; **45**: 1468–88.
68. Lohr KN. Emerging methods in comparative effectiveness and safety: symposium overview and summary. *Med Care* 2007; **45** (10 Suppl 2): S5–8.
69. Green SB, Byar DP. Using observational data from registries to compare treatments: The fallacy of omnimetrics. *Stat Med* 1984; **3**: 361–70.
70. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**: 360–72.
71. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**: 200–8
72. Berger ML, Mamdani M, Atkins D, Johnson ML. 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.
73. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. 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.
74. Macpherson H. Pragmatic clinical trials. *Complement Ther Med* 2004; **12**: 136–40.
75. Relton C, Torgerson D, O'Cathain A, Nicholl J. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. *BMJ* 2010; **340**: c1066.
76. 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.
77. Alford L. On differences between explanatory and pragmatic clinical trials. *New Zeal J Physiotherapy*, March, 2007. http://findarticles.com/p/articles/mi_6848/is_1_35/ai_n28446903/pg_7/?tag=content; col1 (accessed July 2011).
78. Luce BR, Kramer JM, Goodman SN, Connor JT, Tunis S, Whicher D, *et al*. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med* 2009; **151**: 206–9.
79. Strom BL, Eng SM, Faich G, Reynolds RF, D'Agostino RB, Ruskin J, *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.
80. Strom BL, Faich GA, Reynolds RF, Eng SM, D'Agostino RB, Ruskin JN, *et al*. The Ziprasidone observational study of cardiac outcomes (ZODIAC): design and baseline subject characteristics. *J Clin Psychiatry* 2008; **69**: 114–21.
81. 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.
82. Kerridge I, Lowe M, Henry H. Ethics and evidence based medicine. *BMJ* 1998; **316**: 1151–3.
83. Carey TS, Melvin CL, Ranney LM. Extracting key messages from systematic reviews. *J Psychiatr Pract* 2008; **14** (Suppl 1): 28–34.
84. 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.
85. IOM (Institute of Medicine). *Clinical Practice Guidelines We Can Trust*. Washington, DC: The National Academies Press, 2011.
86. Kanouse DE, Kallich JD, Kahan JP. Dissemination of effectiveness and outcomes research. *Health Policy* 1995; **34**: 167–92.
87. Strom BL, Schinnar R, Aberra F, Bilker W, Hennessy S, Leonard CE, *et al*. Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: a randomized controlled trial. *Arch Intern Med* 2010; **170**: 1578–83.
88. Strom BL, Schinnar R, Bilker W, Hennessy S, Leonard CE, Pifer E. Randomized clinical trial of a customized electronic alert requiring an affirmative response compared to a control group receiving a commercial passive CPOE alert: NSAID–warfarin co-

prescribing as a test case. *J Am Med Inform Assoc* 2010; **17**: 411–5.
89. Strom BL, Schinnar R, Jones J, Bilker W, Weiner M, Hennessy S, *et al*. Detecting pregnancy use of non-hormonal category X medications in electronic medical records. *JAMA* 2011, in press.
90. AHRQ. Intent to publish grant and contract solicitations for comparative effectiveness research (CER) projects with funds from the American recovery and reinvestment act (ARRA). *Fed Regist* 2009; **74**: 41913–4.
91. Slutsky J, Atkins D, Chang S, Sharp BA. AHRQ series paper 1: comparing medical interventions: AHRQ and the effective health-care program. *J Clin Epidemiol* 2010; **63**: 481–3.
92. Whitlock EP, Lopez SA, Chang S, Helfand M, Eder M, Floyd N. AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program. *J Clin Epidemiol* 2010; **63**: 491–501.
93. Selker HP, Strom BL, Ford DE, Meltzer DO, Pauker SG, Pincus HA, *et al*. White paper on CTSA consortium role in facilitating comparative effectiveness research: September 23, 2009 CTSA consortium strategic goal committee on comparative effectiveness research. *Clin Transl Sci* 2010; **3**: 29–37.
94. H.R. 3590 The Patient Protection and Affordable Care Act of 2010. Public Law No: 111–148. http://www.govtrack.us/congress/bill.xpd?bill=h111-3590 (accessed July 2011).
95. Policy and Medicine. *Patient Centered Outcomes Research Institute: Names their Board of Governors*, Oct 6, 2010. http://www.policymed.com/2010/10/patient-centered-outcomes-research-institute-names-their-board-of-governors.html (accessed July 2011).
96. Leonard D. *PCORI Board of Governors Begins Its Work*, Sept 2010. http://www.npcnow.org/Public/Newsroom/E-newsletter/2010_e-newsletters/Sept_2010_EVI/PCORI_board_begins_work.aspx (accessed July 2011).
97. Hochman M, McCormick D. Characteristics of published comparative effectiveness studies of medications. *JAMA* 2010; **303**: 951–8.
98. Shepherd J, Rogers G, Anderson R, Main C, Thompson-Coon J, Hartwell D, *et al*. Systematic review and economic analysis of the comparative effectiveness of different inhaled corticosteroids and their usage with long-acting beta2 agonists for the treatment of chronic asthma in children under the age of 12 years. *Health Technol Assess* 2008; **12**: iii–iv, 1–360.
99. Pearson S. From better evidence to better care: using comparative effectiveness research to guide practice and policy. In: *The Hamilton Project. Implementing Comparative Effectiveness Research: Priorities, Methods, and Impact*. Engelberg Center for Health Care Reform at Brookings, June 2009, pp. 55–82.
100. Murray MD. Curricular considerations for pharmaceutical comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2011, in press.
101. VanLare JM, Conway PH, Rowe JW. Building academic health centers' capacity to shape and respond to comparative effectiveness research policy. *Acad Med* 2011; **86**: 689–94.
102. Mitka M. US government kicks off program for comparative effectiveness research. *JAMA* 2010; **304**: 2230–1.
103. Wilensky GR. Cost-effectiveness information: yes, it's important, but keep it separate, please! *Ann Intern Med* 2008; **148**: 967–8.

PART V

Selected Special Methodologic Issues in Pharmacoepidemiology

CHAPTER 33
Assessing Causality of Case Reports of Suspected Adverse Events

Judith K. Jones
The Degge Group, Ltd, Arlington, VA, USA

Introduction

A major component of the evaluation of reports of suspected adverse drug reactions in the clinical setting, or adverse events in a clinical trial, is a judgment about the degree to which any reported event is, in fact, causally associated with one or more suspected drug(s). In reality, a particular event is either associated or is not associated with a particular drug, but the current state of information almost never allows a definitive determination of this dichotomy. Accordingly, a number of approaches to the determination of the probability of a causal drug–event association have evolved over the past several decades. This chapter will first discuss the historical development 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.

Clinical problems to be addressed by pharmacoepidemiologic research

The basic clinical problem to be addressed is illustrated in Figure 33.1. A clinical event occurs within the milieu of a number of possible causal factors. That event either occurred independently or in some way its occurrence was partially or totally linked to one or more of the potential causative agents. The primary task is to determine the degree to which the occurrence of the event is linked to one particular suspected causal agent, in this case a drug or other medicinal agent.

This task of evaluating causality in case reports shares some similarities with the problem of evaluating causality in chronic disease epidemiology, as discussed below and in more detail in Chapter 3. However, in the latter case, causality relates to events in populations and to the assessments of those events in one or more population studies. In individual case reports of suspected adverse reactions to a medicinal product, where data are often incomplete, both the details of the exposure and sometimes the nature of the event make the determination of causality in case reports a major challenge. The evaluator of these cases makes, at the very least, an implicit judgment of causality. Because evaluations of these case reports is such a frequent activity in postmarketing surveillance, it would be optimal to have a coherent, consistent, and reliable method of determining the degree to which there may be a causal relationship between given exposures and specific events. However, there are several attributes of single reports that represent obstacles to such assessments, specifically:

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

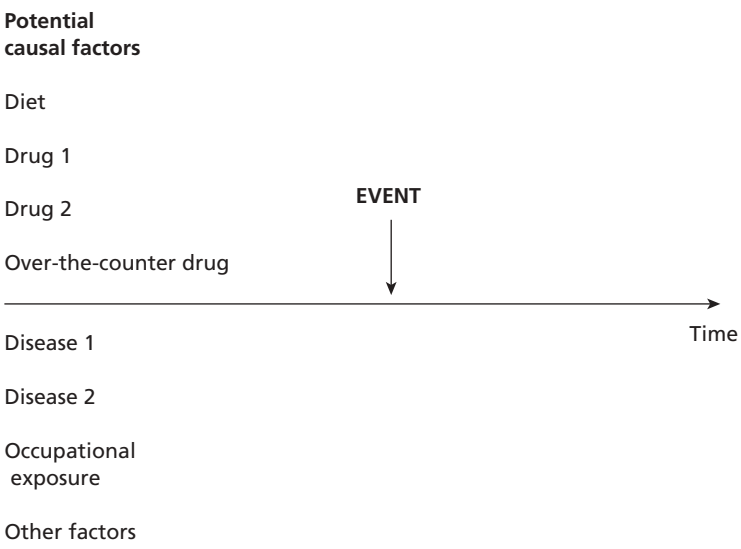


Figure 33.1 Diagram depicting the dilemma for determining causation of an event in a clinical setting. In reality a drug either did or did not cause or contribute to an event. However, given the multiple factors associated with the event, the actual truth can seldom be ascertained. Instead, some expression 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.

1 The usual focus of suspected adverse reaction assessment is an individual clinical event suspected of being associated with exposure to a drug or other medicinal product. The reporting of this event will typically be in the context of a suspicion by the reporter that the event is drug-induced, which will often bias the collection of data required to evaluate other possible causes.

2 The data available about a patient's *drug exposure* in the typical case report are often incomplete, usually missing precise information on prior exposure, duration of use, actual dose ingested, and/or concomitant drugs administered. Such information may be more often reported on events in the hospital setting.

3 The data available on the *adverse event*, including its onset, characteristics, and time course, are also typically incomplete, because the suspicion is usually retrospective and the desired data (e.g., baseline laboratory data) are often not available when the report is made.

4 Data on concomitant diseases and other confounding conditions, such as diet and habits, are typically not available, often because reports are made based upon the specific suspicion of a cause, rather than a differential diagnosis.

Since 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, the challenge has thus been to define general data elements and criteria for determining causality that will apply to most types of suspected adverse reactions. For example, for irreversible events such as birth defects or death, data on de-challenge (the result of discontinuing the drug) and re-challenge (the result of re-introducing the drug) are irrelevant.

Closely linked to the task of determining whether there is a causal relationship between a drug exposure and an event is the *motivation* for making that particular causality assessment and the impact of that inference on any actions taken. 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, for example, continuation of a clinical trial hinges upon the assessment, the reliability of the method becomes more critical. With greater focus on the entire subject of adverse drug reactions and the introduction of concepts of causality assessment into more drug regulatory language, the need for consistent and reliable methods of causality determination[1] has become more important.

Specifically, with the appearance of US Food and Drug Administration (FDA) regulations and draft guidances for reporting in clinical trials of adverse events that are "reasonably" associated with a drug,[2] there is a growing need to describe the basis for defining an association within this setting. Although some of the above-listed uncertainties, such as details of exposure, are less likely to exist in clinical trials, there remain difficulties in assessing the likelihood of association for rare events, which will be considered below. In September 2010, the FDA published Draft Guidances for safety reporting in clinical trials that updates the current requirements in some respects.[3] In March 2003, the FDA proposed to broaden 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, FDA agreed to maintain the former requirement that an event must be "reasonably associated," for it to warrant reporting.[4] These regulations are effective as of March 28, 2011. Use of algorithms in other regulatory environments (e.g., France, EU) are considered below.

Historical perspectives

Development of concepts of causality for adverse reactions

The development of thinking about the causality of adverse reactions has evolved in two disciplines: (i) in epidemiology, and (ii) in the study of individual case reports of adverse reactions. Consideration of both is important.

In the 1950s, epidemiologists grappled with the issue of causality. Yerushalmy and Palmer[5] 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 five criteria to determine the causal nature of an association:

1 the consistency of the association;
2 the strength of the association;
3 the specificity of the association;
4 the temporal relationship of the association; and
5 the coherence, or biological plausibility, of the association.[6]

These criteria continue to be generally used in chronic disease epidemiology, although they have been actively discussed and criticized.[7] They are most appropriately applied to population-based data rather than in the evaluation of individual cases or groups of cases from poorly defined populations. However, in some circumstances where 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 evaluation of collective data on vaccine adverse effects by the Global Advisory Committee on Vaccine Safety of the Immunization Safety Priority Project.[7] Shakir and Layton[8] have cited these criteria as useful for considering the overall data, including spontaneous reports, on an adverse event. Although seldom explicitly noted, the reasoning behind Yerushalmy and Palmer's criteria appeared at about the same time as did thinking about the causal assessment of individual reports of suspected adverse reactions.

Until the last two or three decades, association between a drug and a reported adverse event was typically assumed to be the case if there were a number of similar reports. Considerations of pharmacologic plausibility, dose–response, and timing factors were sometimes implicit, but seldom explicit. This approach is still used in some cases. More often, this tendency has been supplanted by more specific methods, proposed and used in the 1980s, that will be detailed later in this chapter.

The more perplexing single or multiple suspected drug–event associations were typically referred to one or more experts, who generally approached the evaluation by what has been termed "global introspection." In this approach, the experts collected all the facts relevant to the problem at hand, compiled them, and made unstructured judgments to decide the answer. In the causality assessment context, this answer has usually been expressed in terms of a qualitative

probability scale: "definite," "probable," "possible," "doubtful," or "unrelated."[9]

The recognized subjective nature of global reasoning 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 actually likely be drug-associated.[10,11] Shortly thereafter, clinical pharmacologists Karch and Lasagna also recognized the inadequacy of expert "global" evaluations of adverse reactions and developed a decision table, or algorithm, to segment the evaluation of a case into several components.[12] These two groups of investigators identified very similar basic data elements that they felt were necessary for a more standardized assessment:

1 the timing of the event, relative to the drug exposure;
2 the presence or absence of other factors which might also cause the event;
3 the result of withdrawing the drug ("dechallenge");
4 the result of reintroducing the drug ("rechallenge"); and
5 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 derived for chronic disease epidemiology; but in fact, the temporal relationship does also apply. Furthermore, in assessing the causality of either a single report or even a series of cases, outside of a population context, there would be 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 associated with a drug. For example, some patients treated with sulindac developed renal stones containing the drug. Further, in a recent consideration of evaluating single reports, Aronson and Hauben proposed four types of adverse reactions reports where they contend "attribution to the drug is either irrefutable or demonstrable to a high level of confidence."[13] These include:

1 deposition of the drug or metabolite in extracellular or intracellular tissue;
2 a very specific anatomical location or pattern of injury, such as injection site edema;
3 direct tissue injury or physiological dysfunction that can be proven by physicochemical testing, such as esophageal injury with bisphosphonates; and
4 infection resulting from administration of an infectious agent, such as in bacterially contaminated injections.

Following the introduction of these new methods for the assessment of suspected adverse drug reactions, a large number of other approaches were developed,[14–21] either as algorithms, decision tables, or, in at least one case, as a diagrammatic method.[21] These were reviewed and summarized in monographs from two conferences held in the early 1980s on the causality of adverse reactions—one in Morges, Switzerland,[22] and another in Crystal City, Virginia.[23] The vast majority of these methods shared the basic elements originally suggested by Irey and Karch and Lasagna, but many added numerous other details useful for the evaluation of special cases, such as injection site reactions or *in vitro* verification (e.g., Venulet *et al.*[20]). Some included extensive scoring systems linked to relatively extensive algorithms, such as the approach published by Kramer *et al.*[14] A summary of the information categories in the method of Kramer *et al.* is presented in Table 33.1, and selected methods are discussed in detail in the next section of this chapter.

The 1981 Morges conference,[22] the 1983 Crystal City conference,[23] and a 1983 Paris meeting[24] 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 that could be used by regulators and pharmaceutical sponsors

Table 33.1 A summary of the information categories in the method of Kramer *et al.*[14] for determining causality of adverse drug reactions

Axis*	Information category	Number of questions in the axis†	General content
I	Previous experience with drug	4	Literature or labeling information
II	Alternative etiologies	9	Character, frequency of event with disease versus drug
III	Timing of events	4	Timing consistent
IV	Drug levels, evidence of overdose	6	Blood levels, other dose-related events
V	De-challenge	23	All aspects of timing of de-challenge and results
VI	Re-challenge	10	All aspects of re-challenge: circumstances, timing, and results

*Axis in the published algorithm. Although the visual format of the published algorithm appears complex, the axes correspond to the information considered in the majority of causality assessment methods. The authors then weight the answers to the questions to provide a score for each axis which, when summed, gives a numerical estimate of the probability of an association, ranging from 6 or 7 corresponding to "definite" to less than 0 corresponding to "unlikely".

†Each question within an axis relates to a factor that might be considered to contribute to the causality assessment. However, not all questions are asked for any one problem.

alike. An international study group, the Associated Permanent Workshop of Imputologists (APWI) ("imputology" being the French term for causality), was initiated at the Morges meeting and continued into the 1990s.[25,26] Although a consensus method was not established, the Crystal City conference had requested an outside observer (Dr. David Lane, a theoretical statistician) to provide a critique 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 the causality of adverse reactions based on 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 in use elsewhere in medicine, this approach has not been applied to analyses of suspected adverse effects. This method will be discussed in more detail in the next section of this chapter.

After this flux of activity in the mid-1980s, there was more limited activity in the area of adverse event causality, primarily marked by efforts in France in the mid-1990s, where causality assessment was a regulatory requirement in reporting. This resulted in further elaboration of the Bayesian method by Begaud and colleagues in Bordeaux[33] and development of a further scoring method, RUCAM, by Bénichou and Danan.[34] Since this time, although a standard method has not been adopted, 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, in 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, but may be increasing as interest in various methods of analysis of adverse

events has burgeoned. However, causality assessment has been required in France for many years[35] and has been formally considered in a European Community Directive.[36,37] This has resulted in a general consensus on the causality terms used by the European Union member states.[38] Further, since 1994, a formal method of causality assessment for reports of vaccine-associated adverse events has been instituted by Health Canada's Vaccine Safety Surveillance Section, Division of Immunization, Laboratory Center for Disease Control which is conducted by the Advisory Committee on Causality Assessment.[39]

In fact, there are a variety of settings where standard assessments of causality could be useful, 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 reaction should be a common component of the differential diagnosis, and even possibly to the courtroom and the newsroom.[1]

Pharmaceutical manufacturers

Manufacturers of pharmaceuticals and, more recently, biologics (thus combined as biopharmaceuticals) must view causality assessment for events associated with their drugs from the standpoints of both regulatory requirements and product liability. Until recently, US biopharmaceutical manufacturers have not had to consider assessment of causality 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" (earlier, US Code of Federal Regulations 21:310.300; and currently, 21:314.80). The causality assessment did not formally apply to events in clinical trials. However, effective March 28, 2011, new US FDA regulations require causality assessment for determination of reporting certain types of clinical events in clinical trials in the Investigational New Drug (IND) regulations (CFR 21:312.22). The new regulations require the 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. The regulations also include a disclaimer that notes that such a report of a serious unexpected event does not constitute an admission that the drug caused the event. These regulations do not provide criteria or a suggested method; however, they do imply that such methods might be useful.

In 2006, in postmarketing regulations that became effective June 30, 2006, the FDA modified its standard for including postmarketing safety information on the labeling.[40] The new regulatory standard for addition of an event to the product label Warnings section currently provides: "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)). This language was promulgated to replace the previous provision that read "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." (21 C.F.R. §201.57(e)).

Outside of the US, the requirements for manufacturers to consider causality have varied from country to country, but with promulgation of EU directives on pharmacovigilance and related EU activities, the variation may decrease. Many regulatory agencies have requested or implied some type of evaluation to minimize the number of nonspecific events reported.[34] 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.[17,20,21,41–43]

Causality definitely is an issue for pharmaceutical manufacturers in the arena of product liability, especially in the US (see Chapter 9). A number of years ago, Freilich considered many aspects of this,[44] 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.[45,46] 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 regulators

The use by drug regulators of causality assessment of spontaneously reported postmarketing adverse reactions has varied considerably. Most countries' drug regulators have some method of approaching causality, but this method has been most well defined in France, Australia, and certain other countries.[47–49]

In France—owing in part to the considerable original work and interest in adverse reaction causality by a regulator, J. Dangoumou, and his colleagues—all reports of suspected reactions must be evaluated by the "French Method." This method 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."[16,50]

In the US, although a data element for causality was incorporated into the initial file format for the computerized reports of suspected adverse reactions submitted to the FDA, 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.[18,19] This simple, basic method based on only the timing, de- and re-challenge and confounding factors criteria, very specifically excluded the consideration of previous literature reports as a basis for 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, and such a criterion would suppress a signal of a possibly new drug-associated event. The primary use of the assessment by the FDA was administrative, that is 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."

Although the FDA algorithm existed for the reviewers of the reports, the frequency of its actual use was not determined, and this causality data element was removed from the computer file 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). It is of note that Bandekar *et al.*, in 2010, examined and compared key data elements in adverse event reporting forms in ten countries in Europe, North America, Asia, and Africa.[51] Notably, they identified which countries requested data on such items as *re-challenge* and *de-challenge* 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 have applied the Naranjo scoring method.[15] 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 recommending inclusion of it, many published reports do not provide information on confounding drug therapy or medical conditions, data elements considered by those most knowledgeable in adverse reaction assessment to be essential for considering causality. This issue was recognized as one of several problems relating to the publication of adverse reactions in the literature. This problem was first discussed extensively in 1983 at a conference on publication of adverse event reports in the medical literature in Morges, Switzerland. A number of editors of medical publications were present and discussed the quality of information in reported cases. The participants

developed a list of the types of information that would be desirable for published reports, including data that would permit the reader to assess independently the likelihood of the association.[52,53] The conclusion was that publication of case reports should require the specifics of the five elements of the criteria for causality (e.g., details of timing, the nature of the reaction, discontinuation and reintroduction, and alternate causes based on prior history). The need for this publication requirement was underlined 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. Although analysis suggested 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.[54] Nonetheless, even two decades 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 and Drug Safety* and *Drug Safety*.[55,56] Subsequently, the recommendations have been adopted by some other journals, including *Annals of Pharmacotherapy* and *Therapie*.[57]

Methodologic problems to be addressed by pharmacoepidemiologic research

The problem to be solved in determining whether an event is caused by a biopharmaceutical product is to find one or more methods that are reliable, consistent, accurate, and useful for determining the likelihood of association. This problem is compounded by the nature of biopharmaceutical-associated adverse events. They vary in their frequency, their manifestations, their timing relative to exposure, and their mechanism, and mimic almost the entire range of human pathology, as well as adding unique new pathologies (e.g., kidney stones consisting of drug crystals and the oculomucocutaneous syndrome caused by practolol). In addition, since drugs are used to treat illnesses, 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 described, and seldom are the desired baseline and other detailed measurements taken.

The nature of this task, and its context, has generated two divergent philosophies. One philosophy discounts the value or importance of causality assessment of individual reactions, deferring judgment to the results of formal epidemiologic studies or clinical trials.[58] The alternate view contends that the information in single reports can be evaluated to determine at least some degree of association, and that this can be useful, and sometimes critical, when discontinuation of a clinical trial or development of a drug, or, drug withdrawal is a consideration.[59] This latter view has spurred the evolution of causal evaluation from expert consensual opinion based on 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 those that are rare and serious, the question has been raised about whether epidemiologists need to consider using methods for causal evaluations of cases in their formal studies and in clinical trials, since the small numbers available may not be amenable to standard statistical analysis.[60]

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."[61]

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.[9]

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 a cause given an effect, precisely the task of causality assessment.[62] 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.[17,18,19,63,64]

Second, global introspection is uncalibrated. One assessor's "possible" might mean the same thing as another assessor's "probable." This has been well demonstrated in a study of one pharmaceutical company's spontaneous report reviewers, who used both a verbal and numerical scale.[22] These and other shortcomings of global introspection as a causality assessment method for adverse reactions are discussed in detail by Lane, Hutchinson, Kramer, and others.[9,27,64–68] 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 "unassessible/unclassifiable" that essentially represent six levels of global introspection, though they generally incorporate consideration of the more standard criteria for causality.[69] The Portuguese central pharmacovigilance unit (Nucleo de Farmacovigilancia do Centro) utilizes this WHO global introspection method, in part based upon a comparison of results from evaluation of 200 cases by algorithm methods and the WHO global introspection method. They found a relatively moderate to high degree of correspondence of judgments for the reactions more likely associated.[70]

Algorithm/criterial method with verbal judgments

The subsequent attempts to address the limitations of global introspection have resulted in the proliferation of methodologic approaches (see Venulet *et al.*[22] and Herman[23] for reviews and examples of these methods and the appendix in Herman[23] which includes a complete bibliography; also see summaries in Herman and Fourrier[71] and Agbabiaka *et al.*[72]). These methods range from simple flow charts posing ten or fewer questions to lengthy questionnaires containing up to 84 items. However, they share a common basic structure essentially based on the original work by Karch and Lasagna[12] and Irey[10,11]—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 the drug is discontinued (dechallenge), and the response when the drug is subsequently re-administered (re-challenge). 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").

These approaches have advantages when compared to global introspection,[64] since there is a great improvement in the consistency of ratings among reviewers. Because the consideration of each case is segmented into its components (e.g., timing, confounding diseases, etc.), this also allows for a better understanding of areas of disagreement. However, there is still considerable global introspection required to make judgments on the separate

elements of the algorithms or decision tables. These judgments require, in some cases, "yes" or "no" answers where, in fact, a more quantitative estimate of uncertainty would be more appropriate. For example, the reviewer might have to consider whether the appearance of jaundice within 1 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.

This type of approach, with various degrees of complexity, is used by some drug regulatory agencies, such as that of Australia.[48] The FDA algorithm, based on the Irey and Karch and Lasagna concepts, but currently not in official use, was another example of this approach, inquiring sequentially about temporal sequence, de-challenge, re-challenge, 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, since they have neither de-challenge or re-challenge possibilities. To address this, an alternate algorithm for fatal outcome events was developed by Turner in the aftermath of the FDA algorithm.[19]

Algorithms requiring scoring of individual judgments

Many algorithms permit quantitative judgments by requiring the scoring of their criteria. 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,[20] which has now been translated for computer use, to the relatively simpler French method.[16] The method developed by Kramer *et al*.[14] 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, as shown here (Table 33.1[14]). One of the more practical methods of this type was developed by Naranjo, Busto *et al*.[15] This has been adopted in a number of clinical settings and by at least one publisher (*Annals of Pharmacotherapy*) and is shown in Figure 33.2.[15] One of the more recent versions of this type of evaluation was developed by Bénichou and Danan,[34] called the RUCAM method which, like the Naranjo methods, has six criteria with three or four levels of scoring for each criterion to derive an overall score. This has been applied in evaluation of adverse events in HIV clinical trials.[73]

CAUSALITY ASSESSMENT
NARANJO SCORED ALGORITHM

QUESTION	ANSWER Yes	No	Unknown	SCORE
Previous reports?	+1	0	0	_____
Event after drug?	+2	–1	0	_____
Event abate 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 33.2 A critical scored algorithm illustrated by the method of Naranjo *et al*. in wide use.[15] 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. Reproduced from Naranjo *et al*.[15] with permission from *Nature*.

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.[22,47] The spe-

cific manner in which they are used has not been well described in the literature.

Probabilistic methods

Recognition of the various problems inherent in the previously existing methods set the stage for the development of an alternative approach based on the Bayesian probability approach to assessment of causality. This method has provided an opportunity for a fresh look at the issue of causality, and its initial apparent difficulty (due to its requirement for using all available information) raised some new issues about causality assessment of adverse reactions. It has also brought the area of adverse reactions evaluation into a larger discussion of the value of the Bayesian and probabilistic approaches to the analysis of medical and scientific data.[6]

First published as a method for adverse reaction assessment by Auriche,[43] who participated with Lane and others in a working group within the APWI organization, this method was first presented in extensive form in a workshop in 1985 (Figure 33.3).[31] Several examples were published in a monograph and subsequently in early papers.[31,32] The methods have been incorporated into automated versions by both Naranjo and Hutchinson, the latter developing a model using an expert system.[74,75] Naranjo and colleagues have implemented a practical spreadsheet/ automated version called BARDI (Bayesian Adverse Reaction Diagnostic Instrument) and have now applied it to a number of practical adverse event problems.[59,76,77]

The Bayesian method determines 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 33.3. Estimation of this overall probability, the "posterior probability," is based on two components:

1 what is known prior to the event, the "prior probability" which is based on clinical trial and epidemiologic data; and

2 what the likelihoods are, or are not, for drug causation of the components of the specific case, including its history, timing, characteristics, de-challenge and its timing components, re-challenge, and any other factors, such as multiple re-challenges.

The full application of this method requires knowledge of the clinical event, its epidemiology, and relatively specific information about the event's characteristics and kinetics over time. Examples

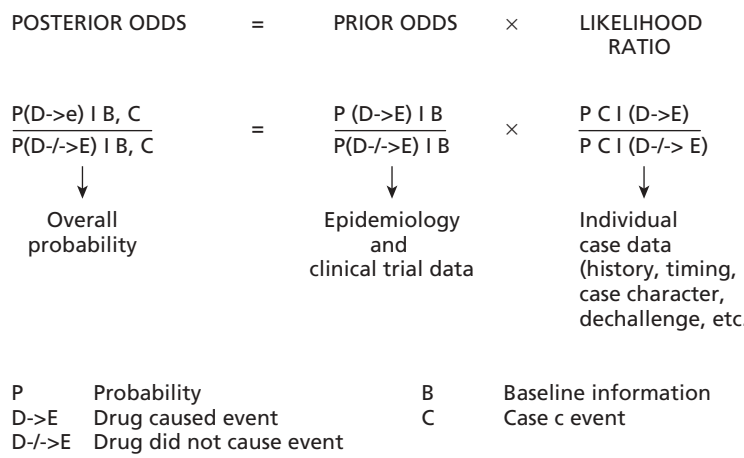


Figure 33.3 The basic equations for the Bayesian analysis of suspected drug-associated events. These provide a structured, yet flexible and explicit approach to estimating the probability that an event is associated with one, or more, drugs, as described in the text and extensive literature dating from Auriche,[43] Lane *et al.*,[31] 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.

have been published for several types of events, including Stevens–Johnson syndrome, renal toxicity, lithium dermatitis, and ampicillin-associated colitis, agranulocytosis, and Guillain–Barré syndrome.[29,59,76] Thus far, this approach appears to be useful for the 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.

With the logistic problem of the length of time required for the actual calculations minimized by automation, the major impediment to more general application of the Bayesian method is the frequent lack of the 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, the 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 now more likely available. So, although this lack of information is sometimes a limitation, it represents both an important challenge and a framework for structuring further understanding. Benichou 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.[78,79]

For this reason, and the fact that more epidemiologic data are becoming available, there appear to be several advantages of using this method for the analysis of suspected drug-associated events:

1 All judgments must be explicit and quantified, which permits 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.

2 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 tenfold 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.

3 Because of the multistep approach to a judgment, combined with a 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 causes.

4 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 serves to 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 will be essential to link these two types of analyses.

An elegant example of the application of this Bayesian method, with an additional complimentary method developed by Begaud and colleagues using the Poisson method for estimating the probability of rare events in populations,[33] has been published by Zapater *et al*.[80] These investigators have nicely 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.

Comparison among the different methods for causality assessment

Several efforts have been made to evaluate and compare these methods. The 1983 conference in Crystal City involved the application of several of

the methods to a standardized case, illustrating a considerable lack of concordance for some methods.[24]

A much more elegant and detailed evaluation of six representative algorithmic methods has been carried out by Pere *et al*.,[81] 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, de-challenge, and alternate etiologies—which tends to underline the lack of considerable information on the events and their characteristics. More recent attempts to quantify agreements on different methods, including global introspection, have been published by Macedo *et al*.[66,67,70]

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. These include:

1 *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?)

2 *The importance of the accuracy of the judgment.* If this evaluation will determine either a specific clinical outcome or, for example, 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.

3 *The number of causality evaluations to be made.* The above considerations must also be weighed against the time required to make judgments on large numbers of reports. This is particularly a dilemma for regulatory agencies and manufacturers, where the need for accurate judgments is pitted against the volume of evaluations to be considered. One approach to this problem is suggested by the FDA's approach to identifying high priority problems according to their newness and seriousness (see Chapter 8).

4 *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, consider a case where 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.[82] In some cases this results from data collection being focused to a much greater degree, as has been initiated in France by Benichou *et al*, where special reporting forms based on disease-specific criteria for events were developed.[82,83] This is also clearly demonstrated in the efforts of Zapater *et al*. in the evaluation of the 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 in the perplexing field of drug-associated hepatotoxicity.[80]

5 *Who will be carrying out the evaluation?* Although no specific studies have been carried out to evaluate the inter-rater 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, which have been described in the preceding sections. Although there was an original hope that there would be some basis for a consensus method,[24] the current state of the field would suggest that this is not likely to be the case, as again evidenced by the ensuing absence of the emergence of a standard method, despite repeated published expressions of need. Several reasons can be suggested. First, a number of individuals and institutions have adopted one or sometimes a few methods and have committed to their use, often

through their choice of data collecting systems or software.[20] Second, the practical aspects of the use of these methods have appeared to play a very real role. Although discussed with excitement as the possible "gold standard" for adverse reaction causality, the Bayesian method was not rapidly embraced, in part because of the difficulty of its use without automation. It was thought that with the lifting of this barrier, and with further use for practical applications, its potential would be realized, but this has generally not been the case. It is likely that the complex Kramer *et al*. algorithm[14] likewise discourages its use in some sectors, although this has not been documented. Again, this is diminished with automation. Third, concern about the misuse of judgment terms or scores within the legal arena has generated concern,[47] particularly given the fact that there is not a gold standard method.

All of these factors suggest the need for considerable further work. This work would appear to fall into several areas:

1 Further definition of the *applications* of causality assessment that is the "output" of the process, so as to better define the desired rigor, accuracy, and usability of the methods. It would appear that there will probably always be needs for simpler and rougher methods, as well as more complete and rigorous methods, when the determination has considerable impact.

2 Further definition of the *critical elements needed* for the evaluation of causality for different types of adverse reactions (e.g., hepatic, hematological, skin, etc.) so that this information may be collected at the time of reporting or publishing a spontaneous event. The need for this event-specific information has long been recognized[16,24,81] 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:

(a) collection of better information on the different drug-associated events, using data collection instruments tailored to the event of interest, and (b) 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[84] and the US NIH) and the regulatory agencies, in particular 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.[85]

3 Gathering of data on these *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.

4 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 method,[81] the Venulet method (J. Venulet, personal communication), and the Bayesian BARDI method.[77] 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.

5 Consideration of *new and different* methods for assessment.

Although it is likely that further work will usually include use of the many available methods, it is of interest that other approaches have emerged. For example, as part of work on patient safety in the US (see also Chapter 45), the methods of "root cause analysis" has emerged to identify the important contributors to adverse events in clinical settings. This approach maps out functional maps of possible contributing factors to not only identify a cause but also determine methods of preventing it. Spath has provided one illustration of this approach.[86] Another approach described by investigators at the University of Toronto, although less generalizable, is the N-of-1 trial that can evaluate

the causality of adverse events in individuals, particularly those who have experienced multiple reactions to drugs.[87]

In conclusion, the topic of causality of adverse reactions continues to represent a challenge. With increased consideration of the need to consider causality as part of the regulatory process, the need for consensus, possibly on more than one method depending on use, continues. One major result of the application of detailed causality assessment, particularly when it is viewed prospectively with collection of data in both pharmacovigilance centers and clinical studies, is that these data can ultimately contribute to the overall need to understand the details of the many drug-associated diseases.

References

1. Jones JK. Uses of drug-event assessment. *Drug Inf J* 1984; **18**: 233–40.
2. US Code of Federal Regulations. Investigational New Drug Procedures, 1987, 21, Part 312.22.
3. 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). *Draft Guidance for Industry: Safety Reporting Requirements for INDs and BA/BE Studies*, September 2010.
4. 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.
5. Yerushalmy J, Palmer CE. On the methodology of investigations of etiologic factors in chronic diseases. *J Chronic Dis* 1959; **10**: 27–40.
6. 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.
7. Anonymous. Causality assessment of adverse events following immunization. World Health Organization Weekly Epidemiological Record, 23 March 2001; **76**: 85–9.
8. Shakir SA, Layton D. Causal association in pharmacovigilance and pharmacoepidemiology: thoughts on the application of the Austin Bradford-Hill criteria. *Drug Safety* 2002; **25**: 467–71.
9. Hutchinson, TA, Lane DA. Assessing methods for causality assessment of suspected adverse drug reactions. *J Clin Epidemiol* 1989; **42**: 5–16.
10. 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.
11. Irey NS. Adverse drug reactions and death: a review of 827 cases. *JAMA* 1976; **236**: 575–8.
12. Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. *Clin Pharmacol Ther* 1977; **21**: 247–54.
13. Aronson JK, Hauben M. Anecdotes that provide definitive evidence. *BMJ* 2006; **333**: 1267–9.
14. 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.
15. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, *et al*. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; **30**: 239–45.
16. Begaud B, Evreux JC, Jouglard J, Lagier G. Unexpected or toxic drug reaction assessment (Imputation). Actualization of the method used in France. *Therapie* 1985; **40**: 111–8.
17. Emanueli A, Sacchetti G. An algorithm for the classification of untoward events in large scale clinical trials. *Agents Actions* 1980; **7**: 318–22.
18. Jones JK. Adverse reactions in the community health setting: Approaches to recognizing, counseling, and reporting. *Fam Community Health* 1982; **5**: 58–67.
19. Turner WM. The Food and Drug Administration algorithm. *Drug Inf J* 1984; **18**: 259–66.
20. 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.
21. Castle WM. Assessment of causality in industrial settings. *Drug Inf J* 1984; **18**: 297–302.
22. Venulet J, Berneker GC, Ciucci AG, eds. *Assessing Causes of Adverse Drug Reactions*. London: Academic Press, 1982.
23. Herman RL, ed. Drug-event associations: perspectives, methods, and users. *Proceedings of Drug Information Association Workshop, Arlington, Virginia*, Oct. 30–Nov. 2, 1983.
24. Jones JK. Drug-event associations: A view of the current status. Epilogue. *Drug Inf J* 1984; **18**: 331–337.

25. Jones JK, Herman RL. Assessing adverse drug experiences (Letter). *Ann Intern Med* 1988; **108**: 160.
26. Venulet J. APWI: past and present. *Drug Inf J* 1991; **25**: 229–33.
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. Begaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? *Brit J Clin Pharmacol* 1994; **38**: 401–4.
34. 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.
35. Moore N, Paux G, Begaud B, Biour M, Loupi E, Boismare F, *et al*. Adverse drug reaction monitoring: doing it the French way. *Lancet* 1985; **2** (8447): 1056–8.
36. CPMP Working Party on Pharmacovigilance. *Procedure for Causality Classification in Pharmacovigilance in the European Community*. EC Document III/3445/91-EN, FINAL, July 1991.
37. Jones JK. Causality assessment of suspected adverse drug reactions: a transatlantic view. *Pharmacoepidemiol Drug Safety* 1992; **1**: 251–60.
38. Miremont G, Haramburu F, Begaud B, Pere JC, Dangoumou J. Adverse drug reactions: physicians' opinions versus a causality assessment method. *Eur J Clin Pharmacol* 1994; **46**: 285–9.
39. Collet J-P, MacDonald N, Cashman N, Pless R, The Advisory Committee on Causality Assessment. Monitoring signals for vaccine safety: the assessment of individual adverse event reports by an expert advisory committee. *Bull World Health Organ* 2000; **78**: 178–85.
40. 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).
41. Stephens M. Assessment of causality in industrial settings. *Drug Inf J* 1984; **18**: 307–14.
42. Stephens MD. *The Detection of New Adverse Drug Reactions*. New York, N.Y.: MacMillan Press, 1985.
43. Auriche M. Approache bayesienne de l'imputabilite des phenomenes indesirables aux medicaments. *Therapie* 1985; **40**: 301–6.
44. Freilich WB. Legal perspectives in causality assessment. *Drug Inf J* 1984; **18**: 211–7.
45. Margulies JB. Epidemiologic causation in the courtroom: Square pegs in round holes? *Drug Inf J* 1991; **25**: 217–27.
46. Henderson TW. Legal aspects of disease clusters. Toxic tort litigation: medical and scientific principles in causation. *Amer J Epidemiol* 1990; **132**: S69–78.
47. Castle W, Baker A. Communication of ADR information: The viewpoint of a British company. *Drug Inf J* 1985; **19**: 375–80.
48. Mashford ML. The Australian method of drug-event assessment. *Drug Inf J* 1984; **18**: 271–3.
49. Wiholm BE. The Swedish drug-event assessment methods. *Drug Inf J* 1984; **18**: 267–9.
50. Dangoumau J, Begaud B, Boisseau A. Les effets indesirables des medicaments. *Presse Med* 1980; **9**: 1607–9.
51. Bandekar MS, Anwikar SR, Kshirsagar NA. Quality check of spontaneous adverse drug reaction reporting forms of different countries. *Pharmacoepidemiol Drug Saf* 2010; **19**: 1181–5.
52. Editorial. Improving reports of adverse drug reactions. *BMJ* 1984; **289**: 898–9.
53. 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.
54. Haramburu F, Begaud, B, Pere JC. Comparison of 500 spontaneous and 500 published reports of adverse reactions. *Eur J Clin Pharmacol* 1990; **39**: 287–8.
55. Kelly WN, Arellano FM, Barnes J, Bergman U, Edwards IR, Fernandez AM, *et al*. Guidelines for submitting adverse event reports for publication. *Drug Safety* 2007; **30**: 367–73
56. Kelly WN, Arellano FM, Barnes J, Bergman U, Edwards IR, Fernandez AM, *et al*. Guidelines for

submitting adverse event reports for publication. *Pharmacoepidemiol Drug Safety* 2007; **16**: 581–7.

57. Kelly W, Arellano F, Barnes J, Bergman U, Edwards R, Fernandez A, *et al*.; International Society of Pharmacoepidemiology; International Society of Pharmacovigilance. Guidelines for submitting adverse event reports for publication. *Therapie* 2009; **64**: 289–94.
58. Louik C, Lacouture PG, Mitchell AA, Kauffman R, Lovejoy FH Jr, Yaffe SJ, *et al*. A study of adverse reaction algorithms in a drug surveillance program. *Clin Pharmacol Ther* 1985; **38**: 183–7.
59. Naranjo CA, Lanctot KL, Lane DA. The Bayesian differential diagnosis of neutropenia and antiarrhythmic agents. *J Clin Pharmacol* 1990; **30**: 1120–7.
60. Jones JK. Epidemiologic perspective on causality assessment for drug associated events. *Drug Inf J* 1986; **20**: 413–22.
61. Agbabiaka TB, Savović J, Ernst E. Methods for causality assessment of adverse drug reactions: a systematic review. *Drug Saf* 2008; **31**: 21–37.
62. Kahneman D, Slovic P, Tversky A. *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982.
63. Koch-Weser J, Sellers EM, Zacest R. The ambiguity of adverse drug reactions. *Eur J Clin Pharmacol* 1977; **11**: 75–8.
64. 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.
65. Kramer MS. Assessing causality of adverse reactions: global introspection and its limitations. *Drug Inf J* 1986; **20**: 433–7.
66. 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**: 697–702.
67. 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**: 885–90.
68. Arimone Y, Bégaud B, Miremont-Salamé G, Fourrier-Réglat A, Moore N, Molimard M, *et al*. Agreement of expert judgment in causality assessment of adverse drug reactions. *Eur J Clin Pharmacol* 2005; **61**: 169–73.
69. World Health Organization Collaborating Centre for Drug Monitoring, Uppsala, Sweden. http://www.who-umc.org (accessed July 2011).
70. 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 Therap* 2003; **28**: 137–43.
71. Herman RL, Fourrier A. *List of References: Third International APWI Meeting*. Paris, November 27–28, 1992.
72. Agbabiaka TB, Savović J, Ernst E. Methods for causality assessment of adverse drug reactions: a systematic review. *Drug Saf* 2008; **31**: 21–37.
73. Bonfanti P, Valsecchi L, Parazzini F, Carradori S, Purterla L, Fortuna P, *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.
74. 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.
75. Hutchinson TA, Dawid AP, Spiegelhalter DJ Cowell RG, Roden S. Computerized aids for probabilistic assessment of drug safety II: an expert system. *Drug Inf J* 1991; **25**: 41–8.
76. Lanctot KL, Naranjo CA. Using microcomputers to simplify the Bayesian causality assessment of adverse drug reactions. *Pharm Med* 1990; **4**: 185–95.
77. 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.
78. Benichou 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.
79. Benichou 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.
80. 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 Safety* 2002; **25**: 735–50.
81. Pere JC, Begaud B, Harambaru F, Albin H. Computerized comparison of six adverse drug reaction assessment procedures. *Clin Pharmacol Ther* 1986; **40**: 451–61.

82. Benichou C. *Guide Practique de Pharmacovigilance*. Paris: Editions Pradel, 1992.
83. Benichou C, Danon G. Experts' opinion in causality assessment: results of consensus meetings. *Drug Inf J* 1991; **25**: 251–5.
84. Lee WM. Drug-induced hepatotoxicity. *N Engl J Med* 2003; **349**: 474–85.
85. Sassolas B, Haddad C, Mockenhaupt M, Dunant A, Liss Y, Bork K, *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**: 60–8.
86. Spath P. Uncover root causes with E&CF charts: Illustrate causal factors in accident sequence. *Hosp Peer Rev* 2000; **25**: 96–8.
87. 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.

CHAPTER 34
Molecular Pharmacoepidemiology

Stephen E. Kimmel,[1] Hubert G. Leufkens,[2] and Timothy R. Rebbeck[3]

[1] Departments of Medicine and of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA

[2] Department of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands

[3] Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA

Introduction

One of the most challenging areas in clinical pharmacology and pharmacoepidemiology is to understand why individuals and groups of individuals respond differently to a specific drug therapy, both in terms of 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 inter-individual 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 42), and last, but not least, the genetic profile of the patient.[1]

Molecular pharmacoepidemiology 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 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 and other biomarkers.

It has been suggested that, on average for each medication, about one out of three treated patients experience beneficial effects, one out of three do not show the intended beneficial effects, 10% experience only side effects, and the rest of the patient population is non-adherent 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 42), 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]

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

Genetic variability in drug response: historical perspective

Although molecular pharmacoepidemiology is a relatively new area of research, the idea that individuals have different susceptibility to 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 (Table 34.1). One of the most well-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 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 genetic 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.

Table 34.1 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 isoniazide *N*-acetyltransferase polymorphism[16]
Inefficacy of codeine as analgesic in poor metabolizers (CYP2D6)[17]

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.

Thanks to numerous human genome initiatives, we also have substantial information about interin-

dividual 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 sequence at one point in the genome inherited either from the father or mother, and the combination of alleles in an individual is denoted 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 (e.g., Cargill *et al.*[18]). It is likely that only a subset (perhaps 50 000–250 000) of the total number of SNPs in the human genome will actually confer small to moderate effects on phenotypes (the biochemical or physiological manifestation of gene expression) that are causally related to disease risk.[19]

However, SNPs are not the only form of genetic variation that may be relevant to human traits and diseases. Copy number variants (CNV) have also been recently identified as another common form of genomic variation that may have a role in disease etiology.[20]

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 define units that remain intact over evolutionary time.[21] These regions define genomic blocks 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 response 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* encompasses not only studies of genetic variability in drug response, but also includes approaches that consider data about thousands of genotypes, as well as responses in gene expression to existing medications.[22,23] 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 are 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 as pharmacoepidemiology to answer questions related to the effects of genes on drug response. Thus, molecular pharmacoepidemiology answers questions related to:

1 the population prevalence of SNPs and other genetic variants;

2 evaluating how these genetic variants alter disease outcomes;

3 assessing the impact of gene–drug and gene–gene interactions on drug response and disease risk; and

4 evaluating the usefulness and impact of genetic tests in populations exposed, or to be exposed, to drugs.

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 raise 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.

Three ways that 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).[5,22–24] 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).[25] 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).[26] 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.[17,27] 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]

Although many drug–CYP2D6 genetic variant interactions have been reported based on experimental or epidemiologic associations, predicting clinical outcomes in daily practice based on such 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.[28] Drug–gene associations shown in one study cannot always be replicated in another one.[29] 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, physiological feedback mechanisms, testing bias), something that is also of critical importance for molecular pharmacoepidemiology.[30]

The genetic polymorphism of thiopurine methyltransferase (TPMT) in treating cancer patients is another example.[10,31,32] In its usual state, TPMT metabolizes thiopurine drugs, which would otherwise be toxic if not excreted. In approximately 90% of individuals, TPMT activity is high and allows normal drug excretion. In 10%, activity is intermediate due to the presence of a heterozygous variant in the TPMT gene. In 0.3%, 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 hematological toxicity. Thus, TPMT genotyping can be determined prior to treatment with these agents to avoid potential toxicities.[33] 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),[34] which has polymorphisms that have been associated with, for example, resistance to antiepileptic drugs.[35] It has been found that patients with drug-resistant epilepsy (approximately one of three patients with epilepsy is a non-responder) 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).[35] 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.[35] 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.

This is well illustrated by the polymorphisms of the β(2)-adrenergic receptor (β(2)-AR), known for their role in affecting response to β-agonists (e.g., albuterol) in asthma patients. In particular, the coding variants at position 16 within the β(2)-AR gene (β(2)-AR-16) have been shown to be important in determining patient response to albuterol treatment.[11] Israel *et al*. showed that the Arg–Arg genotype at β(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 β(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 down-regulation of the receptor) play a critical role. Later clinical and epidemiologic studies, directed at optimizing asthma treatment through β(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.[36]

Pharmacodynamic gene–drug interactions may also result in mixed responses in terms of intended and non-intended 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.[37] 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.[38] 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 B(2) receptor gene are found to be significantly higher in subjects with cough.[38]

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 α-adducin gene and its relation to treatment for hypertension. Cusi *et al*. found a significant association between the α-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.[39] These findings fuelled various pharmacoepidemiologic studies to evaluate whether the α-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.[8,40]

Genetic variability in disease states also can be critical for tailoring drug therapy to patients with a specific genotype related both to 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 over-expression of the HER2 oncogene. The HER2 protein is thought to be a unique target for trastuzumab therapy in patients with this genetically associated over-expression, 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.[41]

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.[42] A relative risk of about 30 in factor V carriers and users of oral contraceptives compared to non-carriers 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.[43] 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.[43] This highlights another important role of molecular pharmacoepidemiology: determining the utility and cost-effectiveness (see also Chapter 38) of genetic screening to guide drug therapy.[44]

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.[23] Depending on the combination of these genotypes, the individual might have a different response in terms of both efficacy and toxicity (see Table 34.2). In the simplified example in Table 34.2, 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.

Table 34.2 Hypothetical response to medications by genetic variants in metabolism and receptor genes

Gene affecting metabolism*	Gene affecting receptor response*	Drug response	
		Efficacy (%)	Toxicity (%)
Wild-type	Wild-type	70	2
Variant	Wild-type	85	20
Wild-type	Variant	20	2
Variant	Variant	35	20

*Wild-type associated with normal metabolism or receptor response and variants associated with reduced metabolism or receptor response.
Modified from Evans and McLeod[23]

A summary of the specific examples cited earlier and their relationship with each of the three mechanisms of genetic variability in drug response is shown in Table 34.3.

Some examples of the progression and application of molecular pharmacoepidemiologic research

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

Table 34.3 Pathways of gene–drug interactions and some relevant examples

Pharmacokinetic	Pharmacodynamic	In pathway of disease
CYP2D6 "poor metabolizer" type and antipsychotic-induced parkinsonism	$ß_2$-adrenergic receptor (ß2AR) and response to $ß_2$-agonists	α-Adducin and salt-sensitive form (diuretic response) of hypertension
Thiopurine methyltransferase defect and toxicity of cancer drugs (e.g., azathioprine)	Dopamine-4 receptor and response to antipsychotics	HER2-overexpression and response to Herceptin®
ABCB1 transporter gene and multidrug resistance in epilepsy (MDR1)	Bradykinin B(2) receptor gene and ACE-induced cough	Factor V Leiden and VTE risk in OC users

form is the cytochrome P450 2C9 variant (CYP2C9).[45–47] Pharmacogenetic studies identified polymorphisms in CYP2C9 that led to altered metabolism of warfarin.[48,49] 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.[50] 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.[51] 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.[52] 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 meta-analysis of studies examining the role of CYP2C9 in warfarin-treated patients demonstrated a significant association between CYP2C9 variants and bleeding risk.[53] Of note, there is still a large amount of interindividual variability in response to warfarin within CYP2C9 genotypes, and numerous studies, including genome-wide association studies (discussed below) have been performed to try to identify other genetic variants that alter warfarin response. From this work, it is now clear 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.[54] 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. This illustrates nicely the complexity of understanding the genetic variability of medication response and the need for increasingly complex molecular pharmacoepidemiologic studies.[55]

The ultimate question that molecular pharmacoepidemiologic studies will have to answer is whether knowing that a patient carries a polymorphism or polymorphisms will lead to better outcomes. The recent development of algorithms to predict a maintenance warfarin dose that combines clinical and genetic data suggests that improvements may be made by incorporating genetic data into dosing algorithms.[56] Further pharmacoepidemiologic studies have demonstrated the validity of these algorithms for predicting warfarin maintenance dose,[57] and a few small randomized trials suggest that genetic-based dosing might improve outcomes (although none to date have been definitive).[58,59] This has led to several, ongoing, large-scale randomized trials comparing warfarin dosing using pharmacogenetic algorithms (that also include clinical factors) to dosing using clinical algorithms.[60–62] The ability to do such studies also emphasizes one of the key advantages to molecular pharmacoepidemiology, the ability to carry out both observational and randomized comparisons of gene-based dosing or drug selection.

Another pertinent example of how pharmacogenetics could lead to molecular pharmacoepidemiologic studies that may guide decision makers in safe prescribing, and also fuel new drug development, is the HIV drug abacavir. Clinical trials have shown that severe hypersensitivity reaction to abacavir (HIV reverse transcriptase inhibitor) is seen in 4% of patients and results in switching to other HIV therapy.[50] These severe reactions could have resulted in cessation of development of the drug. However, the occurrence of hypersensitivity has been linked to the genetic variant HLA B5701, raising the possibility that genetic screening could allow safe use of the drug.[51] Indeed, a series of studies, including randomized trials, have demonstrated the clinical utility of genetic screening prior to dosing abacavir, and this is now standard of care.[63] The abacavir example also shows increasing attention from regulatory authorities for genomic markers. HLA B5701 testing is now included in most labels for abacavir in order to optimize therapy by preventing HLA-related hypersensitivity. Both the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have guidance documents in place in order to implement scientific proof of genotyping into prescribing recommendations.[64,65] Other examples of HLA genotyping include also the prevention of sever skin reaction in ethnic Asian patients using the antiepileptic drug carbamazepine (e.g., HLA-B*1502), also flagged in the product label to consider when prescribing the medicine.[66]

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, 37, and 47).

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.[19,67] 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. This section will begin by highlighting the nature of gene discovery and then 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).

Approaches to gene discovery

While many hypothesize that inherited genetic variation influences the metabolism of drugs and other exposures to influence treatment responses, recent debate has emerged about the nature of this genetic variation. As shown in Figure 34.1, there are a number of categories of genomic variants that could explain the pharmacogenetics of drug metabolism and treatment response. Variants that have large phenotypic effects are likely to be uncommon in the population (A, Figure 34.1). Examples of this class of variants include TPMT and response to chemotherapeutic agents, as discussed earlier in this chapter. In contrast, variants that have large effects and are relatively common (B, Figure 34.1) are unlikely to exist.

More recently, two schools of thought have emerged about the genetic architecture of pharmacogenetics. The first is the common disease–common allele hypothesis,[68] which postulates that commonly occurring inherited variants confer small effects on drug metabolism (C, Figure 34.1). These so-called "low penetrance" alleles have been hypothesized to explain a large proportion of drug response because the attributable risk associated with these variants could be large if the alleles are carried by a large proportion of the population. Accompanying this hypothesis is the notion that the overall disposition of drugs, and the attendant pharmacologic consequences and treatment effects observed in an individual, may result from numerous allelic variants of this type. These variants are typically identified via case–control association studies, including genome-wide association studies (GWAS). GWAS are studies in which randomly selected DNA sequences (selected across the genome to try to identify as much of the variability in DNA as possible) are examined for associations with outcomes, initially irrespective of biological plausibility (discussed in further detail below under Currently Available Solutions).

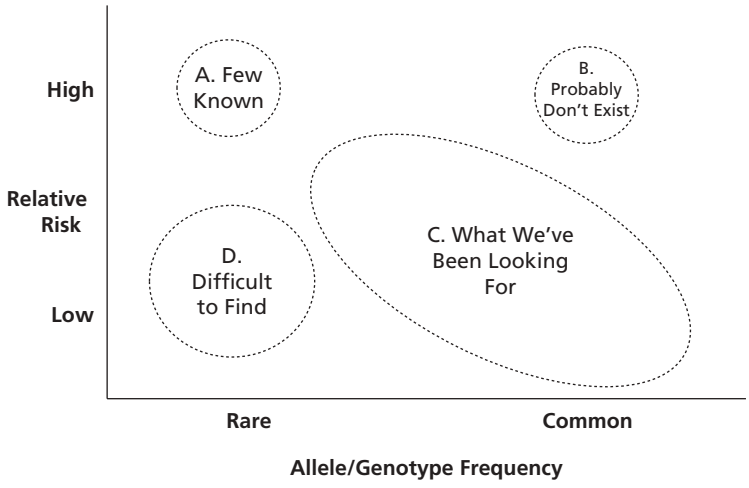


Figure 34.1 Paradigm for genetic effects on drug response and prevalence of genetic variants.

While there are examples of such pharmacogenetic association studies, including GWAS, that have revealed validated associations, there has been limited success in translating these findings into clinical practice. One reason for this limited success can be found from the experience of translating commonly occurring, low penetrance alleles to risk prediction in disease etiology studies. Despite the success in defining many such risk alleles in a wide variety of diseases, few of these have been translated into clinical practice as tools to refine risk assessment, screening, treatment, or other clinically relevant activities. In part, this is due to the small effect sizes of each single risk allele (usually much less than a relative risk of 1.5, which may provide limited clinical utility to most applications), and because combinations of these alleles, if they can be found, confer only clinically relevant effect sizes for extremely rare combinations of alleles having an effect on very limited subsets of the populations.

An additional 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.[69] This "missing heritability" of complex disease suggests that other classes of genetic variation may explain the genetic contribution to common disease.

As shown in Figure 34.1, many have argued that this "missing heritability" in common disease etiology may be explained by a large number of rare variants, each of which may confer very small effects (D, Figure 34.1).[69] To date, there has been little success in confirming the hypothesis that rare variants explain common disease, and essentially no studies of pharmacogenetics. The limited data are due both to the unavailability of cost-effective technologies for whole-genome sequencing or other methods that may allow the detection of these rare variants, as well as statistical methods that allow researchers to identify associations in this setting. The former limitation is likely to be overcome as technology for genetic sequencing becomes cheaper and more available. Methods for identifying associations remain limited, but include: employing linkage disequilibrium in regions of interest to identify collections of rare variants; combining common and rare variants to provide joint information about the effect of variation in a region; studying individuals with extreme phenotypes to target deep sequencing activities; making use of admixture (defined below) and ancestry genomic differences to identify rare variants (discussed further below); using well-annotated families to study inheritance; studying structural variants including deletions and duplications; novel case–control matching strategies that consider not only epidemiologic matching or adjustment algorithms but also matching to specific genomic regions; using pooling strategies to study rare variants; and employing copy number variation (DNA segments that are 1 kilobase or larger and present at variable copy number in comparison with a reference genome).[69]

Despite the debate about whether common low-penetrance variants or rare variants explain disease and pharmacogenetic effects, it seems likely that both classes of these variants, as well as rare variants with large effects, are likely to be responsible for the phenotypic effects of interest. Therefore, hybrid strategies that consider all of these classes of genetic variants must be developed to explain the genetic architecture of common disease and pharmacogenetic response.

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,

Table 34.4 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
	–	c	d		among those with the genotype
–	+	e	f	eh/fg	Effect of medication vs. no medication
	–	g	h		among those without the genotype
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

Source: Modified from Khoury *et al*.[71]

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.[70]

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.[71] The first is as a stratified analysis, comparing the effect of medication exposure versus non-exposure on the outcome in two strata: those with the genetic variant and those without (e.g., see Table 34.4). The second is to present a 2 × 4 table (also shown in Table 34.4). 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 compute both multiplicative and additive interactions.[71] In the example given in Table 34.4, 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 34.4) 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).[72] 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 non-genetic factors, and the interaction between and among these factors.[72–74] One of the reasons cited for non-replication of study findings in molecular pharmacoepidemiology is Type I error.[36] 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.[75] 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.[76] 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,[77] 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).[78] 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 (see Chapter 3) design 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 that are designed to examine the interaction between a genetic polymorphism and a medication will require even larger sample sizes.[79] 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 α-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 non-variant allele, all 1038 participants were available for comparison (Table 34.5). However, when examining the effect of diuretic therapy versus non-use 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 34.4.

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.

Table 34.5 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	A00 103	B00 248	1.0
0	1	A01 85	B01 131	1.56
1	0	A10 94	B10 208	1.09
1	1	A11 41	B11 128	0.77

Case control OR in variant carriers: $OR_{variant} = A_{11}B_{01}/A_{01}B_{11} = 41 \times 131/85 \times 128 = 0.49$
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] with permission.

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.[80] 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 represent admixture of at least three major continental ancestries (African, European, and Native American). Wacholder *et al*.[80] demonstrated that the larger the number of ethnicities involved in an admixed population, the less likely that population stratification can be the explanation for biased associations. Millikan[81] and Wang *et al*.[82] 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*.[83] 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. 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[84] argued that poor study design may be more important than population stratification in conferring bias to association studies. Based on the literature that has evaluated 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.[80] 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

Gene discovery: genome-wide versus candidate gene approaches

There currently are two primary approaches for gene discovery: candidate gene association studies

and genome-wide associations studies (GWAS). In the former, genes are selected for study on the basis of their plausible biological relevance to drug response. In the latter, randomly selected DNA sequences are examined for associations with outcomes, initially irrespective of biological plausibility. GWAS rely on linkage disequilibrium (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* 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 genome-wide association studies: 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.[77] 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.[85] 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.[86] Finally, even if genome-wide scans may identify markers associated with the trait of interest, a challenge will be to identify the causative SNPs.

Clearly, candidate gene and genome-wide approaches are not mutually exclusive. It has been suggested that gene discovery can focus on SNPs or haplotypes based on: (i) strong prior information about biological pathways or linkage data; (ii) information about the functional significance of an SNP or haplotype; and/or (iii) studies that start with a "simple" haplotype involving a small number of SNPs that can be expanded to increase the number of SNPs that constitute haplotypes in a specific region of the genome.[77]

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.[87,88] 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.[71] (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 34.5 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.[87,88] 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).[71]

There are several limitations of the case-only design.[87] 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 the 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 the genotype, by altering response to medications targeted at a specific 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 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 etiological effect. Lohmueller *et al.*[36] observed that many associations are not replicated. This lack of replication can be explained by false-positive reports (e.g., spurious associations), by false-negative reports (e.g., studies that are insufficiently powerful to identify the association), or by 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.*[36] addressed these issues by undertaking a meta-analysis of 25 inconsistent associations and 301 "replication" studies (i.e., by ignoring the initial positive report). Most initial associations were not replicated, but an excess (20%) of replicated associations were 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 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,[89] perhaps due to the "winner's curse" phenomenon which predicts that the initial positive report overestimates the "true" value.[90] 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.

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.

One approach to assessing for possible Type I error is the use of "genomic controls." This approach uses the distribution of test statistics obtained for unlinked markers (genotypes at loci that lie in regions other than the location of the gene of interest) to adjust the usual chi square test for the association of interest. For example, if 20 unlinked markers are studied in addition to the candidate gene of interest, none of these 20 should be associated with disease if they are truly random markers with no biological effect. If one or more of these unlinked markers is associated with disease, this implies that the association represents a Type I

error because associations of these unlinked markers cannot be causally associated with disease, and therefore can only represent false-positive associations. Therefore, the observation of associations with the unlinked markers is a measure of the potential for Type I error. This approach is also useful for assessing for possible population admixture, as discussed below.

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 meta-epidemiologic studies. One example is the International Wafarin Pharmacogenetics Consortium (IWPC). This consortium of over 21 centers across nine countries has worked to combine data from multiple cohort studies in order to develop multiethnic dosing algorithms, attempt to identify uncommon SNPs associated with warfarin response, and perform genome-wide association studies.[91] By combining cohorts, the IWPC now has the largest sample size of any warfarin pharmacogenetics studies.

Another potential solution to minimizing Type II error is through meta-analysis, 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 40). One particularly intriguing approach is the concept of prospective meta-analysis in which studies are planned or identified in advance of performing a meta-analysis so that important elements of study design complement each other across studies and important potential sources of bias that hamper the interpretation of retrospective meta-analyses can be avoided (see Chapter 40). The Clarification of Optimal Anticoagulation through Genetics (COAG) trial[61] and the European Pharmacogenetics of Anticoagulation Trial (EU-PACT),[92] two trials of warfarin pharmacogenetic-based dosing, are doing just that.

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 that use this structure in gene identification.[93,94] 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 future

Without any doubt scientific and clinical developments in biology and (bio)chemistry, particularly in the field of genomics and other biomarkers, have 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 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,[33,56] 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.[38] 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, of drug therapy in an earlier phase.[95] 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 38). Veenstra and colleagues have developed a set of criteria for evaluating the potential clinical and economic benefits of pharmacogenetic testing.[96] 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.

What this all means for the future 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 will be applied in clinical practice, the more drug exposure will be influenced by such tests, making genotype and drug exposure non-independent factors.

Finally, just as for all research, the ethical, legal, and social implications of genetic testing must be considered and addressed.[5,97–99] (See also Chapter 35.) Pharmacogenetic testing raises issues of privacy concerns, access to health-care 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.[99] There also is a concern that medicines will be developed only for the most common, commercially attractive, genotypes, leading to "orphan genotypes."[100,101]

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 Ther* 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**: 660–6.
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, Vos HL, Lemaitre RN, Reiner AP, *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, Boushey HA, Cherniack RM, Chinchilli VM, *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, Fuchs H, Paton V, Bajamonde A, *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, Tans G, Bouma BN, Curvers J, *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. *Br Med J* 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, Sklar P, Ardlie K, Patil N, *et al*. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; **22**: 231–8.
19. Chanock S. Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis Markers* 2001; **17**: 89–98.
20. Barnes MR, Breen G. A short primer on the functional analysis of copy number variation for biomedical scientists. *Methods Mol Biol* 2010; **628**: 119–35.
21. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, *et al*. Linkage disequilibrium in the human genome. *Nature* 2001; **411**: 199–204.
22. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 1999; **286**: 487–91.
23. Evans WE, McLeod LJ. Pharmacogenomics-drug disposition, drug targets, and side effects. *N Engl J Med* 2003; **348**: 538–49.
24. Weinshilboum R. Inheritance and drug response. *N Engl J Med* 2003; **348**: 529–37.
25. Benet LZ, Kroetz DL, Sheiner LB. Pharmacokinetics. The dynamics of drug absorption, distribution, and elimination. In: Hardman JG, Limbird LE, Molinoff PB, Ruddon RW, Goodman Gilman A, eds. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*. New York: McGraw-Hill, 1996, pp. 3–27.
26. Schillevoort I, de Boer A, van der WJ, Steijns LS, Roos RA, Jansen PA, *et al*. Antipsychotic-induced extrapyramidal syndromes and cytochrome P450 2D6 genotype: a case-control study. *Pharmacogenetics* 2002; **12**: 235–40.
27. 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.
28. Borges S, Desta Z, Li L, Skaar TC, Ward BA, Nguyen A, *et al*. Quantitative effect of CYP2D6 genotype and inhibitors on tamoxifen metabolism: implication for optimization of breast cancer treatment. *Clin Pharmacol Therapeut* 2006; **80**: 61–74.
29. Swen JJ, Huizinga TW, Gelderblom H, de Vries EG, Assendelft WJ, Kirchheiner J, *et al*. Translating pharmacogenomics: challenges on the road to the clinic. *PLoS Med* 2007; **4**: e209.
30. Meyer UA. Pharmacogenetics—five decades of therapeutic lessons from genetic diversity. *Nat Rev Genet* 2004; **5**: 669–76.
31. 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.
32. 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 U S A* 1995; **92**: 949–53.
33. Marsh S, McLeod HL. Cancer pharmacogenetics. *Br J Cancer* 2004; **90**: 8–11.
34. Hoffmeyer S, Burk O, von Richter O, Arnold HP, Brockmoller J, Johne A, *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 U S A* 2000; **97**: 3473–8.

35. Siddiqui A, Kerb R, Weale ME, Brinkmann U, Smith A, Goldstein DB, *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.
36. 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.
37. Schafer M, Rujescu D, Giegling I, Guntermann A, Erfurth A, Bondy B, *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.
38. 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.
39. Cusi D, Barlassina C, Azzani T, Casari G, Citterio L, Devoto M, *et al*. Polymorphisms of alpha-adducin and salt sensitivity in patients with essential hypertension. *Lancet* 1997; **349**: 1353–7.
40. van Wieren-de Wijer DB, Maitland-van der Zee AH, de Boer A, Kroon AA, de Leeuw PW, Schiffers P, *et al*. Interaction between the Gly460Trp alpha-adducin gene variant and diuretics on the risk of myocardial infarction. *J Hypertens* 2009; **27**: 61–8.
41. 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.
42. 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.
43. 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.
44. 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**: 569–87.
45. Kaminsky LS, Zhang ZY. Human P450 metabolism of warfarin. *Pharmacol Ther* 1997; **73**: 67–74.
46. Hirsh J, Dalen JE, Anderson DR, Poller L, Bussey H, Ansell J, *et al*. Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range. *Chest* 1998; **114**: 445S–69S.
47. Majerus PW, Broze Jr. GJ, Miletich JP, Tollefsen DM. Anticoagulant, thrombolytic, and antiplatelet drugs. In: Hardman JG, Limbird LE, Molinoff PB, Ruddon RW, Goodman Gilman A, eds. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*. New York: McGraw-Hill, 1996, pp. 1341–59.
48. Takahashi H, Kashima T, Nomoto S, Iwade K, Tainaka H, Shimizu T, *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.
49. Steward DJ, Haining RL, Henne KR, Davis G, Rushmore TH, Trager WF, *et al*. Genetic association between sensitivity to warfarin and expression of CYP2C9*3. *Pharmacogenetics* 1997; **7**: 361–7.
50. 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**: 717–9.
51. 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**: 1816–9.
52. Higashi MK, Veenstra DL, Kondo LM, Wittkowsky AK, Srinouanprachanh SL, Farin FM, *et al*. Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. *JAMA* 2002; **287**: 1690–8.
53. 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.
54. Daly AK, King BP. Pharmacogenetics of oral anticoagulants. *Pharmacogenetics* 2003; **13**: 247–52.
55. Wadelius M, Sorlin K, Wallerman O, Karlsson J, Yue QY, Magnusson PK, *et al*. Warfarin sensitivity related to CYP2C9, CYP3A5, ABCB1 (MDR1) and other factors. *Pharmacogenomics J* 2004; **4**: 40–8.
56. Gage BF, Eby D, Johnson JA, Deych E, Rieder MJ, Ridker PM, *et al*. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin Pharmacol Ther* 2008; **84**: 326–31.

57. Schelleman H, Chen J, Chen Z, Christie J, Newcomb CW, Brensinger CM, *et al*. Dosing algorithms to predict warfarin maintenance dose in Caucasians and African Americans. *Clin Pharmacol Ther* 2008; **84**: 332–9.
58. Caraco Y, Blotnick S, Muszkat M. CYP2C9 Genotype-guided warfarin prescribing enhances the efficacy and safety of anticoagulation: a prospective randomized controlled study. *Clin Pharmacol Ther* 2008; **83**: 460–70.
59. Anderson JL, Horne BD, Stevens SM, Grove AS, Barton S, Nicholas ZP, *et al*. Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. *Circulation* 2007; **116**: 2563–70.
60. van Schie RM, Wadelius MI, Kamali F, Daly AK, Manolopoulos VG, de Boer A, *et al*. Genotype-guided dosing of coumarin derivatives: the European pharmacogenetics of anticoagulant therapy (EU-PACT) trial design. *Pharmacogenomics* 2009; **10**: 1687–95.
61. National Heart, Lung, and Blood Institute. *Clarification of Optimal Anticoagulation through Genetics (COAG)*, 2009. ClinicalTrials.gov identifier: NCT00839657. http://www.clinicaltrials.gov/ct2/show/NCT00839657 (accessed Feb 2010).
62. Washington University School of Medicine. *Genetic Informatics Trial (GIFT) of Warfarin to Prevent DVT*, 2009. ClinicalTrials.gov identifier: NCT01006733. http://clinicaltrials.gov/ct2/show/NCT01006733?term=NCT01006733& (accessed Feb 2010).
63. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, *et al*. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 2008; **358**: 568–79.
64. US Food and Drug Administration (FDA). *Table of Pharmacogenomic Biomarkers in Drug Labels*. http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm (accessed July 2011).
65. European Medicines Agency (EMA). *Note for Guidance on Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories* (EMEA/CHMP/ICH/437986/2006). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002880.pdf (accessed July 2011).
66. Chung WH, Hung SI, Hong HS, Hsih MS, Yang LC, Ho HC, *et al*. Medical genetics: a marker for Stevens-Johnson syndrome. *Nature* 2004; **428**: 486.
67. Nebert DW, Jorge-Nebert L, Vesell ES. Pharmacogenomics and "individualized drug therapy": high expectations and disappointing achievements. *Am J Pharmacogenomics* 2003; **3**: 361–70.
68. Lander ES. The new genomics: global views of biology. *Science* 1996; **274**: 536–9.
69. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al*. Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–53.
70. Kleinbaum DG, Kupper LL, Morgenstern H. Interaction, effect modification, and synergism. In: Kleinbaum DG, Kupper LL, Morgenstern H, eds. *Epidemiologic Research. Principles and Quantitative Methods*. New York: Van Nostrand Reinhold, 1982, pp. 403–18.
71. 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.
72. Kraft P. Multiple Comparisons in studies of gene x gene and gene x environment interaction. *Am J Hum Genet* 2004; **74**: 582–4.
73. Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001; **2**: 91–9.
74. Rao DC. Genetic dissection of complex traits: an overview. *Adv Genet* 2001; **42**: 13–34.
75. 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.
76. Kerwin RW. A perspective on progress in pharmacogenomics. *Am J Pharmacogenomics* 2003; **3**: 371–3.
77. Rebbeck TR, Ambrosone CB, Bell D, Chanock S, Hayes R, Kadlubar F, *et al*. SNP's haplotypes and cancer: applications in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2004; **13**: 681–7.
78. Schreiber W, Stuehlinger HG, Brunner M, Hollenstein U. Preclinical management of thromboembolic disorders. *Semin Thromb Hemost* 1996; **22**: 3–13.
79. Lachenbruch PA. A note on sample size computation for testing interactions. *Stat Med* 1988; **7**: 467–9.
80. 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.

81. 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.
82. 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.
83. 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.
84. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003; **361**: 598–604.
85. Thompson D, Stram D, Goldgar D, Witte JS. Haplotype tagging single nucleotide polymorphisms and association studies. *Hum Hered* 2003; **56**: 48–55.
86. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–9.
87. 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.
88. 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.
89. 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.
90. Capen EC, Clapp RV, Campbell WM. Competitive bidding in high risk situations. *J Petroleum Technol* 1971; **23**: 641–53.
91. Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, Lee MT, *et al*. Estimation of a warfarin dose with clinical and pharmacogenetic data. international warfarin pharmacogenetics consortium. *N Engl J Med* 2009; **360**: 753–64.
92. van Schie RM, Wadelius MI, Kamali F, Daly AK, Manolopoulos VG, de Boer A, *et al*. Genotype-guided dosing of coumarin derivatives: the European pharmacogenetics of anticoagulant therapy (EU-PACT) trial design. *Pharmacogenomics* 2009; **10**: 1687–95.
93. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999; **65**: 220–8.
94. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.
95. Roses AD. Pharmacogenetics in drug discovery and development: a translational perspective. *Nat Rev Drug Discov* 2008; **7**: 807–17.
96. 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.
97. Nuffield Council on Bioethics. http://www.nuffieldbioethics.org/home/ (accessed July 2011).
98. Thomas DC. Genetic epidemiology with a capital "E". *Genet Epidemiol* 2000; **19**: 289–300.
99. Williams-Jones B, Corrigan OP. Rhetoric and hype: where's the "ethics" in pharmacogenomics? *Am J Pharmacogenomics* 2003; **3**: 375–83.
100. Chicurel ME, Dalma-Weiszhausz DD. Microarrays in pharmacogenomics—advances and future promise. *Pharmacogenomics* 2002; **3**: 589–601.
101. Serono VP. *Genomics Raising Disease Redefinition, Other Orphan Issues. The Pink Sheet*, April 10, 2000. http://www.thepinksheet.com (accessed July 2011).

CHAPTER 35
Bioethical Issues in Pharmacoepidemiologic Research

Antoine C. El Khoury,[1] Jason Karlawish,[2] Elizabeth Andrews,[3] and Arthur Caplan[4]

[1] Global Health Outcomes, Merck & Co., Inc West Point, PA, USA

[2] Division of Geriatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[3] Research Triangle Institute Health Solutions, Research Triangle Park, NC, USA

[4] Center for Bioethics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

Research ethics is a discipline that defines the set of norms investigators ought to follow when they conduct research. In the past 50 years as medical research has rapidly evolved, the discipline of research ethics has assumed a largely protectionist posture, principally because of a series of unfortunate scandals and the resulting public outcry.[1–3] As a result, research ethics has focused primarily on protecting human subjects from the risks of research. The goal has been to minimize risks to subjects, rather than minimizing the risks and maximizing the potential benefits for both subjects and society.[4] Themes that run through many of these scandals are scientists' failure to adequately review and disclose research risks and potential benefits, failure to disclose conflicts of interest, and their failure to obtain explicit permission from research subjects. As a result of these events, review of research protocols by an Institutional Review Board (IRB) and strict disclosure of funding sources in addition to informed consent have become the cornerstones for the protection of human subjects from research risks.

Research ethics and research practice have become separate and even, sometimes, antagonistic enterprises. The role society expects of ethics is to regulate science. Current scientific practice reflects this fact. IRB review and the practice of informed consent have become as integral to the design of clinical research as sample size calculations, the accurate measurement of endpoints, or robust statistical analyses.

These and other requirements have been remarkably effective in defining the limits of ethical research, and have made it much less likely that the most egregious ethical errors of the past will be repeated. Overall, they should be viewed as welcome additions to the practice of clinical research. However, serious scientific and ethical problems may arise when the requirements that were developed to guide clinical research are applied to other kinds of research. In particular, standard protections in clinical research are not easily exported and applied to challenges of epidemiologic research. Therefore, as these rules have been applied to pharmacoepidemiologic research, the result has been the parallel development of modifications to the prevailing ethical guidelines and principles, with concomitant increasing consternation and confusion, about how these modifications should be applied beyond clinical settings.

The central problem has been that, while the ethics of human subjects research has been built upon the protection of human subjects, the human

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

subjects involved in many pharmacoepidemiologic studies are quite different. Indeed, it may be difficult to see how the analysis of an existing data set makes the patients whose information contributes to that data set "human subjects" and why this research requires any review by an ethics review board. The idea that a patient can become a subject without his or her knowledge, and without any direct contact with an investigator, is not intuitively clear. Moreover, the risks to the subjects of observational research are not the usual health risks of research that can be balanced against the potential health benefits of research. Harm is not the issue in most pharmacoepidemiologic research. It is almost always what in law and philosophy are referred to as "wrongs," that is, a violation of a person's rights, privacy, or dignity. These risks have been identified among researchers globally.[5–19] While investigators and ethics review boards may be able to balance medical and research risks against medical benefits, they may find balancing these different currencies to be challenging.

In an effort to deal with these problems, investigators, governments, and professional associations have developed regulations and guidelines to provide and disseminate ethical structure to the growing field of epidemiology.[20–24] Most of these guidelines apply equally well to pharmacoepidemiologic research, although this field has begun to develop its own principles.[22–24] These guidelines and regulations have made it clear that the protection of subjects in epidemiologic research represents only one part of the ethical obligations of investigators in epidemiology. Guidelines have addressed four broad categories of ethical issues in epidemiologic research: obligations to society, obligations to funders and employers, obligations to colleagues, and obligations to subjects.[20–24]

Although these guidelines acknowledge a range of ethical obligations, one of these, the investigators' obligations to subjects, has clearly proven to be the most challenging. This is because the procedures of ethical research, like ethics board review and informed consent, may be overly protectionistic or prohibitively difficult in epidemiologic research. Ethical concerns about pharmacoepidemiologic research, and more broadly about epidemiologic research, have therefore focused on the kinds of research that require ethics board review and the kinds of research that require the subject's informed consent.

The answers to these questions define the ethical procedures that allow researchers to have access to information gathered for clinical and administrative purposes. Therefore, investigators face a considerable challenge. They must protect patients' privacy and confidentiality in a way that accomplishes research goals accurately and efficiently. This challenge lies at the heart of the ethics of most pharmacoepidemiologic research.

National and international organizations have created principles that provide a backdrop to the research framework, the most well established being those adopted by the Organization for Economic Cooperation and Development (OECD)[25] in 1980 and more recently by the American College of Epidemiology[24] and the International Society for Pharmacoepidemiology.[23] These recommendations suggest that limits to the collection of data should be sought, especially when it relates to data identifiers, that the quality of data is important, that data use should be specified in advance, and that investigators should adhere to prespecified uses as determined in the protocol. Finally, the OECD suggests a requirement of "openness"—that is, a requirement that goals, uses, and access to data should be a matter of public record, and that individuals should be able to determine whether and how data about them are being used. Despite general agreement about these and other principles, the international community has failed to achieve a consensus about the proper balance of protections and research progress.

A key goal of this chapter is to present an overview of this balance and specifically of the challenges that arise when the principles of research ethics are applied to issues surrounding privacy and confidentiality. In order to accomplish this goal, this chapter will tend to emphasize regulations in the United States (US). This is not because these regulations can or should be generalized to other countries, but simply because at the current time international guidelines vary widely and are often contradictory.[26,27] Therefore, although the experi-

ence of the United States is not universal, these regulations provide a frame of reference for comparison. Where instructive, however, experience from other countries is discussed as well. This chapter also will focus on observational research, which makes up a large proportion of the research in pharmacoepidemiology.

This chapter begins by defining the terms that describe the procedures and requirements of ethical research. These are the normative boundaries in which pharmacoepidemiology must operate in order to maintain the public's trust. If research is to move forward, pharmacoepidemiologists must develop procedures that permit them to balance a need for scientific rigor, on one hand, with the need for ethical requirements, on the other. This chapter will discuss three such strategies and the challenges that investigators face in applying them: delinking subject identifiers from their information, modifications to bioethics board review, and modifications to subject informed consent requirements. This chapter concludes with a critical consideration of some of the available guidelines and regulations, and recommendations for future regulatory efforts.

Clinical problems to be addressed by pharmacoepidemiologic research

The birth and subsequent development of research and scholarship in research ethics, like any field of specialized knowledge, has constructed a language that is particularly its own. This language provides a taxonomy of ethical issues in research and is essential to this discussion because it forms the foundation of any communication and discourse between the fields of ethics and epidemiology. These terms also offer an excellent vantage point from which to examine critically the current emphasis on human subjects' protection and its applicability to pharmacoepidemiologic research.

Pharmacoepidemiologic research

Any productive analysis of the ethics of pharmacoepidemiologic research is critically dependent on a clear and precise understanding of the term "research." Given the frequency with which this term is used by ethicists, investigators, and the public, a definition would seem to be a simple matter. Unfortunately, this has been far from the truth.[28] Yet, perhaps the most well-established definition is also the oldest. In its summary statement (the Belmont Report), the US National Commission for the Protection of Human Subjects defined "research" as any activity designed to "develop or contribute to generalizable knowledge."[29] This is a definition that has been embraced by other scholars, and has become the standard by which a proposed project is assessed.[30]

In clinical research that involves investigators changing the treatment of human subjects, this definition is relatively clear and succinct, although sometimes there may be challenges in categorizing certain clinical studies as research. Unfortunately, in the situation of observational pharmacoepidemiologic research, this definition creates a challenge for researchers and ethics review boards, because it is not always easy to characterize the intent of the person who generates the knowledge and it can be difficult to distinguish between activities conducted as public health surveillance or quality improvement and those conducted as research studies.

The major difficulty arises from the definition itself. What is meant by "generalizable knowledge" and how generalizable should the knowledge be before the study or project is considered research? For instance, data may be gathered as part of a health-care organization's drug surveillance program, the intent of which is to define the patterns of medication use in a local population. This study may generate results that can be used in the local population and the results may be generalized to the local population. However, these same results may not apply to other populations within the same country. Would this study still be considered research? It is not clear, given the definition based on "generalizable knowledge," whether this project should be construed as research, clinical care, or even as a quality improvement activity. Should the data be presented at a professional meeting or submitted for publication, it would most

likely be considered research. These distinctions are important because once a project is identified as "research," investigators must meet a series of requirements designed to protect the patients, who are now human subjects.

This definition of research is particularly problematic in pharmacoepidemiology, because it is often hard to distinguish the routine practice of epidemiology from research. The extremes are evident. The paradigmatic *practice* of epidemiology is public health case finding and surveillance, for adverse drug reactions or drug utilization as examples. This is a social good that we do not, generally, consider being research, although the activities are conducted for the purpose of creating generalizable knowledge upon which to base public health decisions.[31–33] Analogous would be the quality assurance activities of health plans or hospitals, seeking to improve the use of medications in their settings. These sorts of investigations proceed, and sometimes even produce publishable data, without review by ethics review boards. These activities differ from more "research-oriented" epidemiology designed to test hypotheses about drug adverse event associations, drug–drug interactions, drug adherence, or efficacy. These investigations may be identified as research, and they may be required to undergo review by ethics review boards. However, the difference between these two types of activities can be difficult to demarcate. Of particular interest are risk-management programs required by regulatory agencies through which patients can only receive a drug after registering in the program and providing certain medical information. Because these programs are mandatory, they are not considered research and are not reviewed by ethics boards. However, such programs can generate extensive knowledge about the safety and safety use conditions of the product and about the effectiveness and side effects of the risk-management program itself (see Chapter 29). Sometimes such information is not published because the investigators are concerned that journal editors will reject a manuscript because of the lack of ethics board review.[34]

Human subjects

Although it is important that any discussion of research and research ethics be clear about the definition of a research subject, this definition is as elusive as the definition of research, on which it depends. Broadly, though, a useful definition comes from the United States "Common Rule," the set of Federal regulations first promulgated in 1981 that govern research ethics.[35] The Common Rule defines a "research subject" as "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"[35] (US Code of Federal Regulations (CFR) 46.102f). For pharmacoepidemiologists, the key issue here is that the use of information that can be linked to an individual constitutes a contact between an investigator and a human subject. This is true even if the information was gathered in the past and no contact occurs between the investigator and the person. A fundamental issue, then, becomes whether information can be linked to an individual.

This may not be a universally accepted definition. However, the Common Rule applies, at a minimum, to all research carried out by US investigators using Federal funds. In addition, its influence is far greater because the vast majority of institutions that accept these Federal funds have signed an agreement, called a Federal Wide Assurance (FWA), to abide by the Common Rule requirements in all research, regardless of the source of funding.[35] Therefore, the Common Rule serves as *de facto* law governing research at the most productive research institutions in the US and offers a reasonable working definition. Further, even when research is performed outside the United States, if it is done with US Federal support or at an institution with an FWA then it must conform to American regulations governing research ethics.

Privacy and confidentiality

In pharmacoepidemiologic research, the concepts of privacy and confidentiality are of paramount concern. Although they are often discussed

together, they are distinct concepts. It is useful to distinguish them, and to describe individually the ethical basis for requirements of each. Of these, privacy is the most basic and confidentiality is in a sense derivative.

Privacy, in the setting of research, refers to security from unwanted intrusion into physical and personal space including personal information and handling of waste materials from a person. In the case of much epidemiologic research, privacy refers to each individual's right to prevent access to his or her medical records. The right to privacy, and others' corresponding obligation to respect privacy, is justified in part by each individual's right to be left alone.[36] This is a legal way of considering a right to privacy, but privacy has an important social function as well. Viewed in this light, a right to privacy is a precondition for social interaction and cooperation because it allows and requires a degree of trust.[37]

Confidentiality is a derivative right that is based upon the right to privacy. When individuals choose to allow a health-care provider access to personal medical information, they have chosen to waive their right to privacy.[21] Individuals may choose to exercise this right with the expectation, either implicit or explicit, that no one else will have access to that information without the patient's permission. This right to limit the transfer of information, to control the secondary use of information by others, is the right to confidentiality. Like the right to privacy, the right to confidentiality is also based on a basic right to a freedom from interference, in the sense that a right to confidentiality is not possible unless there is an underlying right to privacy. However, the right to confidentiality also engenders a responsibility on the part of the person who has information about another person. The expectation that someone will not disclose the information to a third party creates a fiduciary relationship. That is, it creates an agreement based on a mutually understood set of goals and understandings. This means that confidentiality may be more highly specified by arrangements that may be made at the time that an individual initially grants access to information. For instance, patients may have specific concerns or expectations about ways in which the information they divulge may be used. These expectations may include transfer to a third party in either identifiable or unidentifiable form, or access to particular kinds of information within a medical record, or limits as to the period of time information may be available to others.

The fundamental issue is whether information that was gathered in a clinical setting, where rules of confidentiality apply, can be used for reasons, such as research, that were not part of the conditions of that relationship. Both the law and research regulations are ambiguous over what constitutes a substantive violation of confidentiality. Does the use of records without prior authorization constitute a violation of confidentiality? Or does it constitute a risk of a violation that depends on how those records are used, and on what is done with the information?

In general, society has not articulated clear answers to these questions, in large part because the questions engage well-formed but conflicting political and philosophical views about how society should organize the exchange of information. For example, proponents of communitarianism (the perspective of a community created by voluntary association) argue that the good of the individual is inextricably tied to the public good.[38] Thus, ethical dichotomies that pit individuals against society (such as the unauthorized use of a person's clinical information for research) must be resolved with attention to both personal and public goods.

However, proponents of liberalism, or a rights-based individualism, disagree. From this perspective, what is right exists prior to what is good. This means that any unauthorized use of a person's information threatens to violate a fundamental right to privacy and the potential good derived from that use is not a proper condition to balance against that violation.

In most states in the US, these conflicting views exist in a perhaps deliberately unresolved tension. Laws (or the absence of laws) generally allow procedures that attempt to circumscribe the extremes of either view. Laws are silent on whether medical records can be used for research without the prior

authorization of the patient, although a few states have laws that records can be used for research after Institutional Review Board (IRB) review.[36] Many European nations have very strict protections of individual rights to privacy and confidentiality. For example, Iceland and Sweden have very strict requirements for individual informed consent for the use of identifiable information. Other nations lean toward a more communitarian perspective with respect to epidemiologic research and allow waivers of consent for many studies.

However, US research regulations do provide a set of conditions that permit the use of records regardless of whether the patient authorized their use for research. The key features of these conditions are that the research risks are minimal and the potential violation does not adversely affect subjects' rights and welfare[29] (CFR 46.116). The following sections will discuss both of these key arguments.

Informed consent

Perhaps the most disturbing feature of many of the research scandals in recent history has been the total disregard for informed consent. Informed consent is a legally documented procedure that ensures patients are made aware of all the risks that they may incur and the benefits they may receive when choosing or not choosing a certain therapy or treatment. Every nation which has addressed the subject, as reflected in international codes of ethics and professional society statements about research ethics, recognizes that subjects, or for incompetent patients, their surrogates, are to be told about the nature of research and alternatives to participation, and offered the chance to volunteer to participate or not participate. It is not surprising, therefore, that research ethics guidelines, recommendations, and regulations have stressed the procedural requirement of a subject's informed consent. In order for a subject's consent to be informed, he or she must understand the research and must agree to participate voluntarily, without inducement or coercion.[39]

The regulations governing research informed consent in the US, while not universal, are illustrative of these features[35] (CFR 46.116). The US regulations convey the feature of understanding by requiring that the investigator explain: the research risks, benefits, and alternatives of research participation; the confidentiality of any information obtained; and the procedures for compensation and for contacting a person responsible for the research. Voluntariness is expressed by the requirement that investigators tell subjects that participation in the research study is voluntary, and that subjects have the right to discontinue participation at any time. In some situations (as will be clarified below), informed consent may be modified to be verbal instead of written, or even may not need to be obtained at all. Whether informed consent must always be obtained, and in what form consent should be documented, have been the subject of vigorous debate.[40,41]

Again, while the US guidelines are not universal, they offer a helpful perspective on the complexities that this issue raises. The Common Rule requires that written informed consent be obtained in most research situations[35] (CFR 46.116). However, it makes two notable exceptions. First, written documentation of informed consent is not required if the principal risk of the research is a breach of confidentiality and if the written record is the only link between personal data and the subject's identity[35] (CFR 46.117.c). In this case, whether or not a written informed consent document is used depends on each subject's preferences regarding whether he/she wishes to sign a consent document that could be used to link data with identifiable information. Second, informed consent can be entirely waived if the research meets four conditions[35] (CFR 46.116):

1 the research involves no more than minimal risk to the subjects,

2 the waiver or alteration will not adversely affect the rights and welfare of the subjects,

3 the research could not practicably be carried out without the waiver or alteration, and

4 whenever appropriate, the subjects will be provided with additional pertinent information after participation.

These criteria are often applied to pharmacoepidemiologic research and other forms of research that rely on the use of pre-existing records. The

controversial conditions here are whether the research risks are minimal and whether a waiver of informed consent will adversely affect the subjects' rights and welfare. These are controversial, because in research that involves the use of medical records, the principal risk is the violation of the subjects' confidentiality. A consensus about the proper application of these conditions requires a consensus about whether access to the patient's medical record without the patient's permission is a violation of confidentiality that is greater than minimal risk and violates the subject's rights and welfare.

There are two competing answers to this question. The first relies upon a strict adherence to the principle of respect for autonomy. Accordingly, any unauthorized use of records violates confidentiality, presents more than minimal risk, and adversely affects subjects' rights and welfare. Hence, in all human subjects research, the subject's informed consent could be perceived as an absolute requirement. Although this view follows from strict adherence to some research ethics codes,[42] this is not the view held by most contemporary researchers and ethicists.

Instead, a second interpretation allows for flexibility in the priority of the principle of respect for autonomy. Accordingly, some potential or even actual violations of confidentiality do not adversely affect the subject's rights and welfare or present more than minimal risk. This interpretation requires that we be able to determine to which kinds of information, if any, most people would be willing to grant access. For instance, at one extreme, research using information about patients' sexual activity or certain genetic characteristics might well be perceived as posing a greater than minimal risk. In such a study, obtaining that information without patients' consent might well have an adverse impact upon patients' rights and welfare, depending on the use of the information and the safeguards in place to protect access by third parties. In contrast, information about patients' age and blood pressure might seem to pose only minimal risks, even though blood pressure information could be more predictive of future disability status than the results of a genetic test. Obtaining information without patients' consent must be considered in the appropriate context in a rapidly changing environment, because the potential impact on an individual is heavily dependent on social, economic, and medical factors.

In between these extremes, reasonable people can and do disagree about the magnitude of harm and impact upon rights caused by unauthorized use of information. There are two useful ways to settle this disagreement. The first is to assure that the ethics review board is truly multidisciplinary so that a variety of reasonable views are heard. The second is to require that researchers take steps to minimize the risks and adverse effect upon rights if patient confidentiality is violated. These methods are addressed in the next section.

Minimal risk

Although the general goal of research is to produce knowledge that will benefit society, investigators must also minimize the risks to subjects. It is axiomatic that, as risks to subjects increase, the degree of subject protections, such as ethics review and informed consent, increases as well. The concept of minimal risk attempts to operationalize a risk threshold, above which protections should be stricter. Conversely, subject protections are relaxed if a research protocol does not exceed the level of minimal risk. Although the concept of minimal risk is relatively straightforward, and would apply to most pharmacoepidemiologic protocols, its definition is problematic.

According to US regulations stated in the Common Rule, research risks are "minimal" if "the probability and magnitude of harm or discomfort are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests"[35] (CFR 46.102.i). In most situations, this concept is difficult to operationalize.[43] This is in large part because the definition lacks a clear standard against which to compare the research risks: the daily lives of healthy or "normal" persons, or the daily lives of persons who might be subjects of the research. In pharmacoepidemiologic research where the risk is a potential violation of confidentiality, there is the additional problem

of deciding whether any such violation is ordinarily encountered during daily life, such that a violation in the course of research is "minimal risk."

Ethics review boards

In many countries over the past 30 years, ethics review boards have become central to the practice of research. In the US context, these are committees with at least one "community representative," (discussed in further detail later) appointed by institutions that receive Federal funds to conduct research. In other nations, there are regional or national committees that are appointed by professional organizations or government agencies.

This requirement reflects the consensus that scientists, and science, could benefit from independent review of research protocols. This idea first appeared in the World Medical Association's Declaration of Helsinki in 1964, which requires that an independent committee review all protocols. The Declaration recommends that this committee be responsible for "consideration, comment and guidance," but does not define further the committee's authority to approve or reject protocols that it finds unacceptable.[27] These recommendations have been taken up rapidly, and review boards have become widespread. Their authority has been clarified as well, and these committees typically have the power to review and reject all research that takes place in their institution or in which their institution's investigators are involved. In addition, there have been several independent IRBs established across the US to review protocols for pharmaceutical companies, contract research organizations, and independent researchers. These institutions are generally in compliance with the US Food and Drug Administration (FDA) regulations and adhere to the statutory requirements for protocol reviews as set by Title 21 of the Code for Federal Regulation (CFR 21. 50 and CFR21.56).

In the US, while some states have enacted legislation governing human subjects research, the formal system of review has evolved primarily in a manner that links Federal authority and funding. A committee, referred to as an IRB, is required to review all research that is funded by all Federal government branches that have signed on to the "Common Rule."[35] Examples of Federal agencies that abide by the Common Rule are the National Institutes of Health (NIH), the FDA, and the Agency for Healthcare Research and Quality (AHRQ). Further, as noted above, institutions that have filed an FWA have agreed to abide by the Common Rule requirements in all research, regardless of the source of funding. In most other countries, research regulations are not limited by provisions regarding funding but, instead, apply to all research conducted in that country.

The composition of these review boards varies widely across international boundaries. However, a consistent feature is the need for inclusion of expertise from outside the scientific community. For instance, the US regulations mandate the inclusion of at least one member who is not affiliated with the institution, and one member who may be affiliated but who represents law, ethics, or another non-science discipline[35] (CFR 46.107). Australian regulations mandate a committee's composition by requiring a mix of genders, and by extending the inclusion of non-science representatives.[27] The purpose of these requirements is to introduce accountability to society and minimize conflicts of interest between researchers and scientists who act as research reviewers.

Although review boards have become a commonplace feature on the research landscape, even under US Federal guidelines, not all research requires review. Certain kinds of research can receive expedited review, that is, review by the IRB chair or a designated member of the IRB instead of the full committee, and some may be exempt from IRB review. This is a means to assure that the research risks are truly minor and the research fulfills basic subject protections without expending unnecessary IRB resources. Research that does not require IRB review is any project that does not involve human subjects[35] (CFR 46.101). For example, when investigators use data in which nothing would permit the investigator to identify the individual from whom the data came, ethics board review is not required. In addition, according

to the Common Rule, research may be eligible for expedited review if it poses no more than minimal risk and the research involves "existing data," which means a retrospective analysis of records that exist as of the date the research is proposed[29] (CFR 46.110). Most European nations have similar provisions for expediting the review of research that poses no more than minimal risks to subjects. Internationally, in the past, there was some disagreement about whether pharmacoepidemiologic research should require review. For instance, while the Royal College of Physicians would not require review,[44] the Council for International Organizations of Medical Sciences (CIOMS) recommends ethics board review for all research.[45] The current situation has changed in the UK and there are now specific ethical approval requirements for any study involving people, whether they are physicians, patients, nurses, or care givers in the National Health Service. However, the efficiencies of these requirements have been challenged.[46]

Methodologic problems to be addressed by pharmacoepidemiologic research

There are several procedures available that can protect patient confidentiality. These methods allow patients to control who has access to information. At the time that clinical data are gathered, such as upon enrollment in a health system, a patient can provide a "universal consent" to determine whether his or her medical record can be used for research. This term should not be construed to mean an informed consent to participate in research, because the patient is simply consenting to the generic use of his or her records and not whether to participate in actual protocols. A variation on this method is that patients can shield some aspects of their medical records from use in research. This is possible in some electronic record management systems. For example, patients could place into an electronic "black box," records of certain medications, such as antidepressants. Finally, at the time of the research, patients can be contacted to provide informed consent for the use of their archived records. However, there are two problems in applying these methods to pharmacoepidemiologic research. First, they may not really protect privacy to the degree that investigators and ethics review boards would hope. Second, they may erode the validity of the research findings (as will be further discussed below), and therefore utility for the population that stands to benefit from the research.

First, there is reason for skepticism about whether these procedures (or approaches or strategies) actually foster patient confidentiality. For instance, if individuals must be contacted each time their records may be used in a particular study, the individual may consider such contact intrusive. Furthermore, individuals might consider that their confidentiality has been violated if researchers access research information and contact them directly in order to obtain consent for the use of otherwise de-identified records. Individuals may also refuse participation if contacted for a study they consider irrelevant to their health. An individual may also become alarmed if asked to consent for records to be used in such a study of a disease for which she has not been diagnosed (e.g., a control subject in a case–control study of patients with and without breast cancer). Although these concerns cover very different ground, they all provide grounds for concern that a variety of procedures for protecting privacy may not be ideal.

Validity is a necessary precondition for all ethical research,[47] and research should not be conducted if it cannot answer the hypothesis it claims to test. In pharmacoepidemiologic studies that use archival records, methods that allow patients to control who has access to data can severely limit the validity of the research to be done. For instance, consider the procedure of universal consent, in which each patient is given the opportunity to remove his or her electronic medical records (such as Medicaid data) from use for research. It is certain that at least some patients will opt out. The problem is that willingness to provide consent is generally not random, and varies in ways that may bias study results, as demonstrated in a study of consent bias in the Rochester Epidemiology Project.[48] In an

ethics review board-approved study at the Mayo Clinic, patients were mailed an educational brochure and a request for authorization to allow use of medical records for research. Characteristics of patients who did and did not provide written authorization were compared. Among persons returning the form, the refusal rate was low (3.2%), but the persons declining consent varied from the study population by age, gender, residence, and prior diagnoses, suggesting that the ability to opt out of databases creates a potential bias in the data.[48]

Selective consent by patients may prohibit the evaluation of a key medication–adverse event association if the shielded information is in the pathway between the medication exposure and outcome of interest. For example, the results of a study of a drug–outcome association may be misleading if there is a large increased risk due only to an interaction between the study medication and the confounder drug. The overall study results may show a low-level association between the study drug and the outcome. No interaction could be analyzed. Further, if all patients treated with antidepressants chose to withhold their medical records from any research, the drug–outcome study would show no association, since no data on patients experiencing the reaction would be included in the research data file.

When researchers attempt to contact all patients in a database to seek informed consent, some patients may be unavailable to provide consent because they have died, moved, or changed health plans. Those patients are likely to be distributed in a non-random fashion. The potential bias was demonstrated using data from the Mayo Clinic Rochester Epidemiology Project.[49] Data available from all patients over a 50-year period showed a decrease in the population incidence of hip fracture. Data from only those patients known to be alive and able to give consent would produce results showing an increasing risk of hip fracture over time. This consent issue poses particular challenges in studies requiring long periods of exposure or follow-up, studies evaluating events of long latency, and the evaluation of intergenerational effects of medications.

The number of studies using archival records will likely increase with the growing availability of electronic records and increasing interest in answering important drug safety questions. However, as the number of studies increases, there will undoubtedly be a decreasing consent rate if all studies require consent. Jacobsen *et al*. showed a high consent rate among persons who returned consent forms, but only 79.3% of persons contacted returned the consent forms.[48] Another study of consent, a drug safety study within a population of members of a Minnesota health plan, showed much poorer participation. In this study, with more representative results, only 19% of individuals contacted provided consent, and only 52% provided any response.[50]

An additional problem is encountered in the conduct of large, multi-institutional case–control studies in which access to a large amount of data must be reviewed in order to identify the cases and controls prior to contacting the appropriate patients for consent. Ethics review boards typically waive the requirement for consent in the initial case-finding review of records, and evaluate the consent used when patients are invited to participate in the study. Applying the current Common Rule framework to these studies requires separate review by ethics review boards from each participating institution of the same protocol. Issues raised by these ethics review boards and encountered in the review process may relate less to true local differences in the research environment than to the administrative differences of each institution's ethics review board process. In the absence of a more streamlined approach to the current ethics review board process, the time and cost of seeking multiple approvals discourage the conduct of these studies that may have important public health implications.

These issues are even more complex when attempting to conduct studies across multiple countries in which ethics requirements, consent procedures, and data confidentiality regulations may differ substantially. Harmonizing study procedures to accommodate differences in cultural norms, medical care practice patterns, and regulatory oversight while maintaining the integrity of the research

design requires significant involvement of individuals at the local level.

Currently available solutions

These methodologic challenges pose considerable obstacles to the conduct of pharmacoepidemiologic research. For records-based studies using data not directly identifying subjects, investigators have relied on the confidentiality policies governing the use of information in the individual institution. For studies using identifiable records, investigators receive guidance and direction, if they receive it at all, through a process of negotiation with local ethics review boards, whose task is to balance the requirements of the research design with the rights and welfare of prospective subjects. Because the tension between ethics requirements and the exigencies of pharmacoepidemiologic research require this balancing process, in a very real sense the ethics of pharmacoepidemiologic research is a negotiated agreement between investigators and one or more review boards. The available solutions to the methodologic challenges outlined in the previous section, therefore, depend upon two factors. First, they depend upon the steps investigators can take in gathering and handling data. Second, they depend upon the degree to which review boards can and should be involved in research, and on their ability to review research in a manner that is both competent and efficient. We examine each of these in turn.

The past several years have seen a rapid movement toward legislative protections for data privacy both in the US and internationally. These legislative approaches to protect the confidentiality of medical data provide potentially strong protections and safeguards on the creation and reuse of confidential information. For instance, the European Union (EU) Directive that went into effect October, 1998 covers all information that is either directly identifiable or information from which identity could be inferred.[51] The EU Directive requires first, consent for all uses of information beyond that for which the information was originally collected. Safeguards on the use and transfer of information are required as well. Each institution must have a data controller/data privacy officer, who is accountable for appropriate procedures and use of data within the institution. In addition, data cannot be transferred from a member state of the EU to another country outside the EU unless that country has safeguards at least as stringent as those of the EU. Notably, however, member states may grant deviations from some provisions of the Directive for activities of substantial public interest. Interestingly, there is no mention of ethics review boards in the Directive. All research would presumably: (i) be conducted with explicit consent, (ii) be conducted only with delinked records, or (iii) be exempted by a specific member state as a type of activity of substantial public interest.

For pharmacoepidemiology, a number of implications of the Directive are of concern. For example, pharmacovigilance activities currently must be conducted using identifiable data. A requirement for patient consent would stifle the collection of a substantial proportion of cases and therefore hinder the ability to identify signals of drug safety problems. Furthermore, analysis of secondary information (from clinical trials or administrative databases) for research questions not anticipated at the time patients signed consent would not be possible without additional consent. Very little research could be conducted using secondary files from which direct patient identifiers have been deleted. This restriction is due to the broad definition in the Directive of identifiable and "indirectly identifiable" data.

In the US, the Health Insurance Portability and Accountability Act (HIPAA) of 1996 called for Congress to pass legislation on medical data privacy, and for the Department of Health and Human Services to promulgate regulations if Congress failed to act. While Congress considered numerous bills that promised stricter scrutiny of research and tighter protections, none was passed. Therefore, the Privacy Rule (Standard for Privacy of Individually Identifiable Health Information, in Title 45 of the Code of Federal Regulation, Part 160 and Subparts A and E of Part 164) was developed, and went into effect April 14, 2003 (see www.hhs.gov/ocr/hipaa). The Privacy Rule

offers greater protections of privacy, restrictions on the uses to which existing data can be put, and requirements that individuals must be able to determine who and why others may have access to their personal data in many cases outside of standard medical practice. The rule applies to "covered entities" or organizations that generate and manage personally identifiable health data. While some researchers may not be directly covered by the rule, they generally must obtain access to information from organizations considered covered entities.

Of specific interest for pharmacoepidemiologists are the strategies for protecting confidentiality and enabling researchers to access existing data sets. Under the rule, data sets that are de-identified can be disclosed and used freely. The Privacy Rule defines de-identified data as: (i) a data set from which 18 specified items that could be considered identifiers have been removed, and (ii) a data set which the covered entity knows will not be used alone or with other data for the purpose of subsequently identifying individuals. The covered entity can alternatively use a satisfactory statistical method to de-identify the data set while maintaining some of the 18 elements.

However, epidemiologists would rarely find a data set stripped of these 18 elements appropriate for research because the elements include some items that are essential for research. For example, any specific date field related to an individual would have to be removed. Specific dates are usually required to evaluate sequence and timing of drug exposures and adverse events.

There are several methods researchers can use to gain access to a data set that has not been completely de-identified. First, patient authorization can be obtained. Second, the requirement for patient authorization can be waived by either an IRB or a Privacy Board (which is defined in the rule) if certain conditions are met, such as limits on access to the data, and assurances that the research could not be conducted without the waiver. Third, a "limited data set," that contains some of the 18 elements considered identifiers (e.g., dates and geocodes) can be provided to a researcher if a "data use agreement" has been signed by the researcher assuring the appropriate use and disclosure of the information for research.

There are two additional features of the Privacy Rule that are important in protecting research. A data set can be considered to be de-identified even though a covered entity maintains a code by which the de-identified database can be relinked to personally identifiable data. The code itself cannot be disclosed. In some early drafts of legislation and rules, retention of these codes would not have been allowed. The ability to relink a data set to original data in order to supplement a de-identified data set with information on risk factors, outcomes, or extended follow-up time can be critically important in pharmacoepidemiologic studies. In addition, the Privacy Rule has preserved access by researchers to patient information in certain circumstances for activities "preparatory for research." For example, a preliminary review of medical records is often important to identify patients potentially eligible for a study prior to approaching a patient for consent. Researchers may have access under the Privacy Rule only if the identifiable data are necessary for the preparatory work, and the identifiable information may not be removed from the covered entity as it is reviewed. However, early implementation of the Privacy Rule suggests that the intended balance between protecting patient confidentiality and promoting careful research has not yet been realized, as some covered entities are reluctant to permit research access to data even when all aspects of the Privacy Rule are honored. More recently, the US Institute of Medicine requested the Committee on Health Services to evaluate the impact of the HIPAA privacy rule on the conduct of health related research and to generate recommendations accordingly. The committee concluded that the HIPAA privacy rule is not protecting privacy of patients effectively and is affecting negatively the conduct of meaningful health-related research.[52] The committee recommended the department of Health and Human Services to develop a new framework for patients' privacy to increase security protection and distinguish between ethical review requirements for information-based research versus interventional research.[52]

There are also opportunities to improve and maybe standardize the ethics board review process. Ethics review varies widely from country to country, and, depending on each country's culture and needs, there may be differences between countries or even within one country. However, standardizing the basic ethics board review processes may be more efficient and effective especially when researchers are involved in multicountry pharmacoepidemiologic studies. In existing guidelines there is general agreement that protocol review by ethics review boards is valuable in principle. However, there is considerably less agreement about what kinds of pharmacoepidemiologic research require this review.

In some cases, it is not even the features of the research, but the source of funding that determines whether ethics board review is necessary. For example, as noted previously, the Common Rule regulations apply only to research that is conducted using Federal funds or research that is conducted in institutions that have agreed to follow these regulations voluntarily. The result is that while some researchers are required to apply for ethics review board approval, other researchers whose research presents the same kinds of research risks are not. Although this distinction on the basis of funding source respects the limits of Federal authority in intrastate activities, it lacks moral force.

As examples of efficient protection of human subjects, the Common Rule[35] and the International Society for Pharmacoepidemiology (ISPE)[23] positions seem the most sensible. This means that investigators and ethics review boards will at times need to negotiate the kinds of research that achieve standards such as "existing data" and minimal risk. However, this negotiation is a far better system to assure adequate subject protection for research than a system in which decisions are either entirely left in the hands of the investigators or made by others.

Nevertheless, this system of research oversight, and its heavy dependence on ethics board review, means that oversight can vary widely among institutions. This variability creates enormous administrative challenges for pharmacoepidemiologic investigators, challenges that may be magnified in the case of multicenter research that crosses international borders, including developing countries. Certainly, sensitivity to local issues may be a desirable feature for the ethical review of research, particularly if institutions have special populations or circumstances that warrant special scrutiny of protocols. However, this variability may also be the result of variability in the quality of the ethics review board's skills and resources.

The ability of ethics review boards to review research in a manner that is both competent and efficient addresses issues of the training and certification of membership and resources for handling the volume of new and renewing research protocols. In general, the requirements for the skills and knowledge needed for ethics review board membership are handled by the local ethics review board. No certification exists to assure that ethics review board members possess adequate understanding of research ethics and regulation. Finally, ethics review boards are funded through indirect means, such as the general pool of indirect funds generated from grants. Potential ways to improve the quality and efficiency of ethics board review include training and certification of board members, reduction in the amount of paperwork for routine monitoring of protocols, and explicit funding that is proportionate to an ethics review board's workload.

The future

The variability and quality of ethics board review pose significant challenges for pharmacoepidemiologic investigators. These should be the focus of future efforts to harmonize research regulations and set minimum standards for ethics review board competency and funding. However, these solutions do not adequately address a larger problem. Although ethics review boards may offer a reasonable procedural solution to ethics review, it is less clear how ethics review boards should make the sorts of decisions that are required of them. Specifically, it is not clear how ethics review

boards and investigators should balance ethical and methodologic requirements. Without a careful consideration of this balancing process, any efforts at regulation, and particularly efforts to standardize ethics board review and boost their resources, will achieve only limited success.

The idea of balancing is not new. Traditional approaches to balancing the ethical and methodologic requirements of research typically use as their guide the research risks. In most guidelines, and the Common Rule is an excellent case in point, increasing risks to study subjects requires increasing attention by full ethics board review and the informed consent process, including written documentation of informed consent.[31] Seen in this light, there is a simple proportional relationship between research risks and subject protections such as informed consent. This relationship describes the degree of subject protections solely in terms of the balance of the risks and potential benefits to the subjects of the proposed research.

The problem, though, is that this relationship is too simple for the situation of pharmacoepidemiologic research. The ethical requirements of traditional biomedical research do not fit entirely with the practice of pharmacoepidemiologic research. The risks to the subjects of most epidemiologic research are not the usual health risks of research that can be balanced against the potential health benefits of research. They are instead largely risks of another kind. The chief risk is the violation of confidentiality, which is really a civil, rather than a medical, risk.

We suggest that investigators and ethics review boards should consider an additional factor in this relationship: the value of the knowledge to be gained[35] (CFR 46.111a). An ethical justification for this position begins, first, with the example of social services research. United States research regulations currently include an exception for studies designed to evaluate social programs[35] (CFR 46.101). The implicit argument for this exception is that these social programs offer clear and evident value. They contribute in an important way to the social good. Studies designed to evaluate them, even if these studies bear all of the markings of "research," are considered to be exempt from the requirements of ethics board review and subject to informed consent that govern the ethical conduct of research. In a sense, the requirements of ethical research are suspended for studies that offer significant and generally agreed upon value.

This is an extreme case of balancing value against research risks. Indeed, it effectively removes research involving social programs from the purview of ethics oversight. This example is informative not only because it is so extreme, but also because studies of social programs have a great deal in common with pharmacoepidemiologic research. Pharmacoepidemiology's goals of studying medication use and identifying adverse drug reactions are directed as much toward the preservation of the public's health as they are toward the production of generalizable knowledge. The value of pharmacoepidemiologic research is therefore as clear and as readily evident as it is in studies designed to evaluate social programs. On these grounds alone, a compelling argument might be made that some kinds of pharmacoepidemiologic projects, like projects to evaluate important social programs, should be exempt from research review.

Of course, this argument may not be equally cogent and convincing for all pharmacoepidemiologic research because pharmacoepidemiologic research, like any research, spans a continuum. Certainly studies of adverse drug reactions resemble closely the example of social program research. This is one standard, perhaps the highest standard, for a study's potential to produce valuable knowledge. Phrased somewhat differently, the knowledge must be immediately relevant and applicable to the subjects who are being studied. In pharmacoepidemiologic research, one example might be a study of adverse drug reactions among individuals taking a certain medication. Results of this research would have immediate consequences for the health of the patients, or "subjects," for whom data are gathered.

Other studies may be done for private companies or organizations following rigorous methodologic standards but where the findings would not be made public or shared with anyone outside the sponsoring organization. It is difficult to know how

to balance concerns for privacy against the desire of private entities to obtain pharmacoepidemiologic data. Studies like these should arguably be held to a different ethical standard because they do not hold the immediate possibility of clinically relevant knowledge that could be applicable to the people involved. The problem is that no public and national body exists to decide what kinds of research achieve this level of value.

The central ethical issue in pharmacoepidemiologic research is deciding what kinds of projects will generate generalizable knowledge that is widely available and highly valued, and do this in a manner that protects individuals' right to privacy and confidentiality. The problem is that these two ends differ in kind. The knowledge generated by pharmacoepidemiology is health-related knowledge about such things as the risks and benefits of medicines. In contrast, individuals' right to privacy is a matter of civil law. Although the two are frequently cast as in need of balancing, it may not be possible to weigh a certain amount of knowledge to be gained against a certain amount of confidentiality and privacy to be lost.

Instead, perhaps the most productive approach will be to determine what kinds of procedures and practices warrant crossing thresholds of confidentiality and privacy in the pursuit of valuable knowledge. Such a discussion should include research, but should not by any means be limited to it. For example, society allows journalists wide access to gather and disseminate information, provided the journalist adheres to standards of practice (such as preserving the confidentiality of sources) and journalism is still viewed as a valuable instrument for preserving a democratic society.

Therefore, if the ethical requirement of informed consent is absolute and inviolable, then any balancing would be indefensible. However, this is not a tenable solution, nor is it a solution that would be consistent with the way that society responds to a need for valuable information in other settings. Further public discussion is needed to identify ways in which the policies and procedures for the protection of privacy and the maintenance of confidentiality are fair and consistent with the requirements imposed on other sectors of society.

Acknowledgement

The authors would like to thank David Casarett for his contribution to the chapter which was published in the previous edition of this book.

References

1. Advisory Committee on Human Radiation Experiments. *Final Report*, no. 00000848-9, vol. 061. Washington, DC: Government Printing Office, 1995.
2. Caplan AL. Twenty years after. The legacy of the Tuskegee Syphilis Study. When evil intrudes. *Hastings Cent Rep* 1992; **22**: 29–32.
3. Beecher HK. Ethics and clinical research. *N Engl J Med* 1966; **274**: 1354–60.
4. Kahn JP, Mastroianni AC, Sugarman J. *Beyond Consent. Seeking Justice in Research*. New York: Oxford University Press, 1998.
5. Duszynki KM, Beilby JJ, Marley JE, Walker DC, Pratt NL. Privacy considerations in the context of an Australian observational databse. *Pharmacoepidemiol Drug Saf* 2001; **10**: 587–94.
6. Verstraete AG, Stichele RV, Deliens LHJ. Ethical issues in pharmacoepidemiology in Belgium. *Pharmacoepidemiol Drug Saf* 2001; **10**: 595–9.
7. De Castro MS, Goldmin JR, Ferreira MBC. Ethical issues in a Brazilian hospital. *Pharmacoepidemiol Drug Saf* 2001; **10**: 601–5.
8. Rawson NSB. Ethical issues in pharmacoepidemiologic research using Saskatchewan administrative health care utilization data. *Pharmacoepidemiol Drug Saf* 2001; **10**: 607–12.
9. Hallas J. Conducting pharmacoepidemiologic research in Denmark. *Pharmacoepidemiol Drug Saf* 2001; **10**: 619–23.
10. Rathmann W. Data safety and drug safety in Germany: a closing gap? *Pharmacoepidemiol Drug Saf* 2001; **10**: 625–30
11. Tzimis L. Ethical issues arising in pharmacoepidemiologic research in a community hospital in Crete, Greece. *Pharmacoepidemiol Drug Saf* 2001; **10**: 631–33.
12. Czeizel AE. The ethical issues of the postmarketing surveillance of drug teratogenicity in Hungary. *Pharmacoepidemiol Drug Saf* 2001; **10**:635–9.
13. Barbui CB. Drug epidemiology in Italy: methods and ethical issues. The example of psychiatry. *Pharmacoepidemiol Drug Saf* 2001; **10**: 647–52.

14. Feely J, Williams D. Pharmacoepidemiology—an Irish perspective. *Pharmacoepidemiol Drug Saf* 2001; **10**: 641–5.
15. Tanaka K, Morita Y, Kawabe E, Kubota K. Drug use investigation and prescription monitoring in Japan. *Pharmacoepidemiol Drug Saf* 2001; **10**: 653–8.
16. Coulter DM. Privacy issues and the monitoring of sumatriptan in the New Zealand intensive medicines. *Pharmacoepidemiol Drug Saf* 2001; **10**: 663–7.
17. Evans JMM, McNaughton D, Donnan PT, MacDonald TM. Pharmacoepidemiological research at the medicines monitoring unit, Scotland: data protection and confidentiality. *Pharmacoepidemiol Drug Saf* 2001; **10**: 669–73.
18. Primozic S, Kos M, Mrhar A, Rvnik I. Systemic conditions for performance of pharmacoepidemiologic studies in Slovenia. *Pharmacoepidemiol Drug Saf* 2001; **10**: 675–8.
19. Kaeser Y, Bruppacher R. Ethical and legal considerations in a Swiss study on the quality of pharmacotherapy in the city of Basel. *Pharmacoepidemiol Drug Saf* 2001; **10**: 685–8.
20. Last JM. Guidelines on ethics for epidemiologists. *Int J Epidemiol* 1990; **16**: 226–9.
21. Beauchamp TL, Cook RR, Fayerweather WE, Raabe GK, Thar WE, Cowles SR, et al. Ethical guidelines for epidemiologists. *J Clin Epid* 1991; **44**: 151S–69S.
22. Andrews E. Data privacy, medical record confidentiality, and research in the interest of the public health. *Pharmacoepidemiol Drug Saf* 1999; **8**: 247–60.
23. Epstein M. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**: 200–8.
24. McKeown RE, Weed DL, Kahn JP, Stoto MA. American College of Epidemiology Ethics Guidelines: Foundations and Dissemination. *Science Engineering Ethics* 2003; **9**: 207–14
25. Organization for Economic Cooperation and Development. *Guidelines on the Protection of Privacy and Transborder Flow of Personal Data*. Recommendation of the OECD Council, September 1980.
26. Mann RD, Bertelsmann A. The issue of data privacy and confidentiality in Europe—1998. *Pharmacoepidemiology Drug Saf* 1999; **8**: 261–4.
27. Brody BA. *The Ethics of Biomedical Research: An International Perspective*. New York: Oxford University Press, 1998.
28. Lynn J, Johnson J, Levine RJ. The ethical conduct of health services research: a case study of 55 institutions' applications to the SUPPORT project. *Clin Res* 1994; **42**: 3–10.
29. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: US Government Printing Office, 1979.
30. Levine RJ. *Ethics and Regulation of Clinical Research*, 2nd edn. Baltimore, MD: Urban and Schwartzenberg, 1986.
31. Gostin LO. *Public Health Practice vs. Research: A Report for Public Health Practitioners Including Cases and Guidance for Making Distinctions*, May 24, 2004. http://www.cste.org/pdffiles/newpdffiles/cstephresrpthodgefinal.5.24.04.pdf (accessed April 2010).
32. Center for Disease Control and Prevention. *Distinguishing Public Health Research and Public Health Nonresearch*, July 2010. http://www.cdc.gov/od/science/integrity/docs/cdc-policy-distinguishing-public-health-research-nonresearch.pdf (accessed July 2011).
33. Department of Health and Human Services. *Quality Improvement Activities—FAQs*. http://answers.hhs.gov/ohrp/categories/1569 (accessed July 2011).
34. Andrews EB, Kaye J, Van Bennekom C. The REMS publication paradox. *Oncology* 2009; **23**: 715–17.
35. *Department of Health and Human Services, Protection of Human Subjects*. Title 45 Part 46: Revised. Code of Federal Regulation; June 18, 1991.
36. Warren S, Brandeis L. The right to privacy. *Harv Law Rev* 1890; **4**: 193–220.
37. Fried C. Privacy: a rational context. *Yale Law J* 1968; **77**: 475–93.
38. Avineri S, de-Shalit A. *Individualism and Communitarianism*. New York: Oxford University Press, 1992.
39. Grisso T, Appelbaum PS. *Assessing Competence to Consent to Treatment*. New York: Oxford University Press, 1998.
40. Brett A, Grodin M. Ethical aspects of human experimentation in health services research. *JAMA* 1991; **265**: 1854–7.
41. Truog RD, Robinson W, Randolph A. Is informed consent always necessary for randomized, controlled trials? *N Engl J Med* 1999; **340**: 804–7.
42. The Nuremberg Code. Reprinted in: Brody, H. *The Ethics of Biomedical Research. An International Perspective*. New York: Oxford University Press, 1998, p. 213.
43. Freedman B, Fuks A, Weijer C. In loco parentis. Minimal risk as the ethical threshold for research upon children. *Hastings Cent Rep* 1993; **23**: 13–19.
44. Royal College of Physicians. Research involving patients. Reprinted in: Brody, H. *The Ethics of Biomedical*

Research. An International Perspective. New York: Oxford University Press, 1990, pp. 315–20.

45. Council for International Organizations of Medical Sciences. International Guidelines for ethical review of epidemiological studies. Reprinted in: Brody, H. *The Ethics of Biomedical Research. An International Perspective*. New York: Oxford University Press, 1998, pp. 225–32.
46. Hunter D. Efficiency and the proposed reforms to the NHS research ethics system. *J Med Ethics* 2007; **33**: 651–4.
47. Freedman B. Scientific value and validity as ethical requirements for research: a proposed explication. *IRB* 1987; **9**: 7–10.
48. Jacobsen SJ, Xia Z, Campion ME, Darby CH, Plevak MF, Seltman KD, Melton LJ 3rd. Potential effect of authorization bias on medical record research. *Mayo Clin Proc* 1999; **74**: 330–8.
49. Melton LJ. The threat to medical-records research. *N Engl J Med* 1997; **337**: 1466–70.
50. McCarthy DB, Shatin D, Drinkard CR. Medical records and privacy: empirical effects of legislation. *Health Serv Res* 1999; **34**: 417–25.
51. Olsen J, Breart G, Feskens E, Grabauskas V, Noah N, Olsen J, *et al*. Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Int J Epidemiol* 1995; **24**: 462–3.
52. Institute of Medicine. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Report Brief*, February 2009.

CHAPTER 36
The Use of Randomized Controlled Trials for Pharmacoepidemiologic Studies

Samuel M. Lesko[1] *and Allen A. Mitchell*[2]
[1] Northeast Regional Cancer Institute, Scranton, PA, USA
[2] Slone Epidemiology Center at Boston University, Boston, MA, USA

Introduction

When properly conducted, randomized controlled trials (RCTs) are considered the gold standard for demonstrating the effectiveness of a new medication because they provide unbiased estimates of effect. While RCTs are generally used to evaluate beneficial drug effects, the advantages of this study design also make it ideal for obtaining an unbiased estimate of the risk of adverse outcomes.

During the premarketing phases of drug development, RCTs involve highly selected subjects and in the aggregate typically 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. However, 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). Identification and quantification of these potentially important risks require large studies, which typically are conducted after a drug has been marketed. Because of design complexity and costs, large controlled trials have not generally been considered in the pharmacoepidemiologist's armamentarium for the postmarketing evaluation of drugs. However, in evaluating the best method to assess the risk of serious but rare adverse reactions to pediatric ibuprofen, the authors adapted this approach.[1] That experience is the basis for this chapter and may serve as a guide to the use of randomized trials for the postmarketing assessment of drug safety.

Clinical problems to be addressed by pharmacoepidemiologic research

Pharmacoepidemiologic methods are classically used to quantify risks and benefits of medications that could not be adequately evaluated in studies performed during the premarketing phase of drug testing. In this chapter, the authors will consider the role of postmarketing randomized trials in assessing only the risks of medications; however, the same principles can be applied to the postmarketing evaluation of the benefits of medications.

As noted above, premarketing studies are typically too small to detect modest differences in the incidence rates (e.g., relative risks of 2.0 or less) for common adverse events or even large differences in the incidence rates for rare events, such as those that affect one per 1000 treated patients (see

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

Chapters 1 and 4). Modest differences in risk of non-life threatening adverse events can be of substantial public health importance, particularly if the medication is used by large numbers of patients. For example, following the introduction of angiotensin converting enzyme (ACE) inhibitor use in patients with congestive heart failure, case reports of severe hypotension began to appear in the literature. Although similar events were noted after initial use of other medications (e.g., vasodilators) in congestive heart failure patients, reliable estimates of the risk for different classes of medications were not available. Because of the high prevalence of the indication, differences in risk too small to be detected by conventional RCTs were judged to be clinically important, and a large RCT was conducted to resolve the question.[2]

Modest risks are especially relevant for nutritional supplements or drugs available, or being considered for, over-the-counter (OTC) sale, because these agents are likely to be very commonly used and are widely viewed by the public as safe. 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). However, 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 and then control for their effects. To do so requires the relevant information to be included in the data to be analyzed, but information on important confounding factors is frequently incomplete or unavailable. Although surrogate variables are often used (e.g., years of education to reflect socioeconomic status), these may be poor measures of the underlying confounding factor, and their control therefore may not eliminate confounding.

An investigator observing a crude (i.e., unadjusted) association between a drug and an effect attempts to control for confounding by adjusting for one or more factors. If a crude odds ratio (or relative risk) of 5.0 (for example) remains essentially unchanged after all known confounders have been controlled, residual confounding is usually not considered an important concern; although the true (unconfounded) odds ratio may be somewhat smaller than the unadjusted estimate, it is generally assumed to be of similar magnitude. On the other hand, in the same example, if the adjusted odds ratio (or relative risk) is closer to the null value of 1.0, there is empirical evidence of confounding in the data, and the adjusted odds ratio is usually considered the "best" (least biased) measure of the association. However, it is not possible to determine whether there remains any residual confounding in this best estimate, which if *completely* controlled might reveal that the true association is still weaker or even non-existent.

We have direct experience with this concern. Infants treated in newborn intensive care units frequently receive medications and intravenous fluids through indwelling catheters, and low doses of heparin are often infused to help maintain the patency of these catheters. In 1986, we published the results of a case–control study of the use of intravenous heparin in relation to the risk of intraventricular hemorrhage (IVH) in low-birth-weight infants.[3] In this study, 66 infants with IVH (cases) were compared to 254 infants with no evidence of IVH (controls), matched on study hospital and duration of observation. Compared to no heparin exposure, the matched odds ratio for heparin exposure on the day prior to detection of IVH was 14 (95% confidence interval [CI]: 5.4–34). As additional potential confounders were taken into account, the magnitude of the association became progressively smaller (Table 36.1). Adjustment by logistic regression for the matching factors, birth weight, volume of parenteral fluids administered, and the presence of pneumothorax reduced the odds ratio to 3.9 (95% CI: 1.4–11), which did not

Table 36.1 Effect of potential confounding factors on the relation between heparin exposure and intraventricular hemorrhage in 320 premature infants

Model	Potential confounders included	Odds ratio*	95% Confidence interval
1	Hospital and duration of observation	14.	5.4–34
2	Model 1 + birth weight	7.5	2.8–20
3	Model 2 + IV fluids	4.4	1.6–12
4	Model 3 + pneumothorax	3.9	1.4–11

*Calculated by logistic regression controlling for the potential confounders listed.
Reproduced from Lesko *et al*.[3] with permission from Massachusetts Medical Society.

change further when other potential confounding factors were added to the multivariate model. Although we described an observation that was statistically significant, biologically plausible, and clinically important, we concluded that control of confounding may have been incomplete and that, ". . . the association could have been partly, or even wholly, due to the severity of the infants' underlying conditions rather than to the use of heparin." We suggested that the question could only be answered by a randomized trial. A second observational study also found an increased risk (odds ratio 1.96) among infants who received doses of heparin above the lowest quartile of exposure.[4] However, neither study could adequately control for confounding. Uncertainty about the association persisted until 1997, when results were published from a randomized, double blind trial of heparin added to umbilical catheters used to treat premature infants.[5] Among the 113 study infants, Chang, *et al.* found no difference in the incidence of intraventricular hemorrhage between the heparin treated and control groups ($p = 0.6$). Although the odds ratios from the earlier observational studies were moderately large (3.9 and 1.96) and statistically significant, these "best" estimates of risk were likely due to confounding by one or more factors not completely controlled in the analyses.

Another, perhaps more familiar, example is the purported cardioprotective effect of estrogens. Because women have a lower incidence of heart disease than men, and incidence increases substantially after menopause, it has been hypothesized that estrogens reduce the risk of heart disease. Although not all published observational studies have confirmed this association,[6,7] numerous studies, including at least one large prospective cohort study, have reported significantly lower risks of cardiovascular events among women using postmenopausal hormone replacement therapy.[8,9] The Women's Health Initiative (WHI) was a large, complex clinical investigation of several strategies intended to prevent cardiovascular disease and cancer in postmenopausal women.[10] The study included a placebo-controlled, randomized clinical trial of hormone replacement therapy and the risk of coronary heart disease. This component of the WHI was closed early when a planned interim analysis indicated that the risk of coronary heart disease was significantly elevated among women randomized to hormone replacement therapy (hazard ratio 1.29; 95% CI: 1.02–1.63).[10] Further, hazard ratios were also elevated for breast cancer (1.26; 95% CI: 1.00–1.59), stroke (1.41; 95% CI: 1.07–1.85), and pulmonary embolism (2.13; 95% CI: 1.39–3.25). It seems likely that incomplete control of confounding (and perhaps measurement error as well) in the observational studies obscured these associations. As these examples demonstrate, when residual confounding is a possible explanation for an apparent association, serious consideration needs to be given to whether the hypothesis needs to be tested in a non-observational design.

Weak associations deserve particular attention. Although there are important exceptions, the general view is that the stronger the association,

the more likely 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." This qualification is no more appropriate than for studies reporting weak associations. As a consequence, even after rigorous efforts have been made to control for confounding, many experienced epidemiologists consider small relative risk estimates to be most compatible with no association, regardless of the confidence interval (or *p* value). Whether or not one subscribes to this view, it is advisable to use extreme caution in making causal inferences from small relative risks derived from observational studies.

When there is a basis for concern about residual confounding, one may wish to consider using a non-observational study design. The authors faced just this situation when they considered how to best assess the safety of pediatric ibuprofen. Ibuprofen is a non-steroidal anti-inflammatory drug (NSAID) that has become widely used among adults in the US, first by prescription and then as an 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 would likely be widely used for the treatment of fever, which is typically a minor and self-limited condition. (Whether and when it is appropriate to treat fever in children will not be considered here.) Given the generally benign nature of this indication, it is reasonable to require greater assurance of safety than may 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 Food and Drug Administration required additional data concerning the risk of rare but serious adverse events before it would approve pediatric ibuprofen for OTC sale.

What approach would best provide this information? Observational postmarketing studies, especially case–control studies, are one source of data for very rare conditions. However, the circumstances surrounding ibuprofen use in 1989–1990 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.[1] 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*. It was apparent, then, that to provide a valid assessment of the risks of pediatric ibuprofen, a study must be able to distinguish the risks of the drug from the risks associated with the illness for which ibuprofen was given.

Methodologic problems to be solved by pharmacoepidemiologic research

The phenomenon described above for pediatric ibuprofen is known as confounding by indication (also referred to as indication bias, channeling, confounding by severity, or contraindication bias; see also Chapters 37 and 47). 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."[11] 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, however, this is not easily done (see Chapter 37).

Confounding by indication is a particular concern in a number of settings. When there is a single therapy for an illness, and all patients receive the therapy (i.e., are "channeled" to the treatment), it is not possible to control for confounding in an observational study simply because no patients are left untreated to serve as controls. For example, it is standard practice to administer artificial surfactant to premature infants at risk for respiratory distress syndrome of the newborn. If any infants are not treated, they are likely to differ from treated infants in that they may have a very mild form of the illness, or they may have a major congenital malformation and not be expected to survive. Thus, they are also likely to have different risks for many clinical outcomes. While it may be rare for all patients with a given illness to be treated in exactly the same way, this situation is not unusual for subgroups of patients. For example, all patients with diabetes are not treated with insulin, but patients with type I (insulin-dependent) diabetes are. In general, observational studies are most informative when patients receiving different medications are similar with respect to their risks of adverse events. Cohort studies will be compromised if there is no reasonable alternative to the study treatment, including no treatment, to serve as a control. Case–control studies may be infeasible if one cannot identify controls that, aside from any effect of the exposure, are equally at risk of having the outcome diagnosed as the cases.

When there is at least one alternate treatment option and it is possible to control for obvious confounding, observational studies can contribute to our understanding of a medication's risks, particularly where the adjusted relative risk is large. However, as discussed above, a small relative risk (e.g., 1.3) can easily be an artifact of confounding by an unknown factor or by incomplete control of a recognized confounder.

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 an RCT. There is nothing inherent in an RCT that precludes a pharmacoepidemiologist from designing and carrying out these studies. To the contrary, the special skills of a pharmacoepidemiologist can be very useful in performing large-scale RCTs after a drug is marketed.

Overview of classic randomized controlled trials

RCTs are most commonly used during the premarketing phases of drug development to demonstrate a drug's efficacy (and to gather general information concerning safety). By randomization, one hopes to make the distributions of confounding factors (both known and unknown) equal in all groups. If the study is sufficiently large, the assigned treatment is the most likely explanation for any observed differences in the clinical outcomes (improvement in the illness or the occurrence of adverse events) between the treatment groups. By definition, participants in observational studies are not assigned treatment at random, where the choice of treatment may be determined by the stage or severity of the illness or by the patient's poor response to or adverse experience with alternative therapies, which can introduce bias.

Sample size

In homogeneous populations, balanced treatment groups can be achieved with relatively small study sizes. In heterogeneous populations (e.g., children less than 12 years of age), a large sample size may be required to insure the equal distribution of uncommon confounders between study groups (e.g., infants versus toddlers versus school-age children). Study size is determined by the need to assure balance between treatment groups and the magnitude of the effect to be detected. Large randomized studies minimize the chance that the treatment groups are different with respect to potential confounders and permit the detection of small differences in outcomes.

Blinding

Blinding is used to minimize detection bias, and is particularly important where the outcome is at all subjective. Reporting of subjective symptoms by study participants and the detection of even objectively defined outcome events may be influenced by knowledge of the medications the patient is using. For example, if a patient complains of abdominal pain, a physician may be more likely to perform a test for occult blood in the stool if that patient was being treated with ibuprofen rather than acetaminophen. Thus, follow-up data collection will only be unbiased if all parties (patients, health-care providers, and investigators) are unaware of the treatment assigned. Blinding may not be possible for non-drug treatments such as diet, exercise, and surgery, and double blinding (i.e., of patients and health-care providers or investigators) may be difficult to achieve and maintain in drug studies as well, particularly if either the study or control medication produces specific symptoms (i.e., side effects) or easily observable physiologic effects (e.g., change in pulse rate or blood pressure).

Choice of control treatment

The hypothesis being tested determines the choice of control treatment. Placebo controls are most useful for making comparisons with untreated disease, but may not represent standard of care and have been challenged as unethical.[12] Further, it may be difficult to maintain blinding in placebo-controlled studies, as noted above. Studies employing an active control typically utilize common drug treatments, which frequently represent the standard of care. Although often considered more ethical and easier to keep blinded because the illness and symptoms are not left untreated, these studies do not permit comparison with the natural history of the illness.

Data collection

Data collection in even a small premarketing clinical trial is generally resource intensive. Detailed descriptive and clinical data are collected at enrollment, and extensive clinical and laboratory data are collected at regular and often frequent intervals

during follow-up. In addition to the data needed to test the hypothesis of a clinical benefit, premarketing trials of medications must also assess safety and therefore must collect extensive data on symptoms, physical signs, and laboratory evaluations, which adds to the high cost of these trials. In contrast, data collection in a large simple trial (LST) can be quite simple (as discussed below).

Data analysis

While data collection in classic clinical trials is onerous and expensive, analysis of the primary hypothesis tends to be straightforward and involves a comparison of some measure of the outcome event in different groups. Analyses involving repeated measures, subgroups of study subjects, or adjustment to control for incomplete or ineffective randomization may be performed, but they add only modest complexity. In contrast, data analyses in observational studies tend to be quite complex because of the need to adjust for potential confounders.

Despite the strengths of randomization and relative ease of analysis, several features of the classic RCT limit its use as a postmarketing study design. First, the complexity and cost of traditional premarket RCTs, with their detailed observations and resource-intensive follow-up, make very large studies of this type generally infeasible. Second, it may be unethical to conduct a study in which patients are randomly assigned a potentially harmful treatment. However, if the study can be simplified and use the epidemiologist's tools to track patients and collect follow-up data, it may be possible to both control costs and make a large study feasible. The ethical dilemma can be resolved by studying only questions that are truly important to the public's health and for which the answers are not known.

Generalizability of results

The usual clinical trial conducted during the premarketing evaluation of a drug almost always involves highly selected patients, often in settings that differ markedly from the "real world"; as a consequence, the results may not be generalizable to the large numbers of patients who may use the medication after licensing. One of the attractive features of observational studies is that they tend to reflect the real world experience of medication use and clinical outcomes, and their modest costs permit studying large numbers of patients.

Currently available solutions

Large simple trials

Large, simple trials (LSTs) may be the best solution when it is not possible to completely control confounding by means other than randomization. If the volume and complexity of data collection can be kept to a minimum, there is no reason that large trials cannot be conducted. Indeed, the US Salk vaccine trial of the early 1950s is an example of a very large trial.[13] More recently, large randomized trials have been used to test the efficacy of therapeutic interventions, especially in cardiology,[14–17] or to evaluate dietary supplements or pharmaceuticals for primary prevention of cardiovascular disease and cancer.[18–23] This approach has also been used successfully to evaluate the risk of adverse drug effects when the more common observational designs have been judged inadequate.[2,24,25] 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.

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.[26] 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.[27] 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 of interest to pharmacoepidemiologists.

Power/sample size

Study power is not simply a function of the number of subjects enrolled. It 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, and duration of observation (or follow-up). Power requirements can be satisfied by studying a population at high risk, enrolling a large sample size, 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 long latency periods may be best studied with long duration follow-up. Selecting high-risk populations can be problematic: while an elderly population may meet criteria for high risk (e.g., gastrointestinal bleeding or cardiovascular events), a study limited to this group may lack generalizability and 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. Epidemiologists 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 by mailed questionnaires, telephone interviews, or online using a secure website. Because the study will be limited to clear and objective outcomes (see below) which can be confirmed by medical record review or other means, self-report by study participants 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 maintenance organization where it is likely that important outcomes will by 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 large simplified randomized trial appropriate?

LSTs are appropriate when all of the conditions in Table 36.2 apply.

Important research question

Although a simple trial will 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

Table 36.2 Conditions appropriate for the conduct of a large randomized trial

1.	The research question is important.
2.	Genuine uncertainty exists about the likely results.
3a.	The absolute risk is small and confounding by indication is likely
or	
3b.	The relative risk is small, regardless of the absolute risk.
4.	Important effect modification (interaction) is unlikely.

there is a clear need for a reliable answer to a question concerning the risk of a serious outcome. A minor medication side effect, such as headache or nausea, may not be trivial for the individual patient but may not warrant the expense of a large study. However, if the question involves the risk of premature death, permanent disability, hospitalization, or other serious events, the cost may well be justified.

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.[28] 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). We support this principle and would extend its use to evaluate when it is appropriate to conduct an LST to test a hypothesis related to the risk of an adverse clinical event. Very large randomized 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. For example, the antiestrogen tamoxifen may improve breast cancer survival, but may do so only at the cost of an increased risk of endometrial cancer. Appropriately, a randomized trial was undertaken to resolve uncertainty in this situation.[18]

Power and confounding

LSTs will only be needed if (i) the *absolute* risk of the study outcome is small and there are concerns about confounding by indication, *or* (ii) the *relative* risk is small (in which case, there are inherent concerns about residual confounding).[17] 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 studies would be appropriate.

No interaction between treatment and outcome

An additional requirement for LSTs is that important interactions between the treatment and patient characteristics (effect modification) are unlikely.[17] In other words, the available evidence should suggest that the association will be qualitatively similar in all patient subgroups. Variation in the strength of the association is acceptable among subgroups, but there should be no suggestion that the effect would be completely reversed in one or more subgroups. Because of the limited data available in a truly simple trial, it may not be possible to test whether an interaction has occurred, and the data collected may not be sufficient to identify relevant subgroups. Because randomization only controls confounding for comparisons made between the groups that were randomized, subsets of these groups may not be strictly comparable with respect to one or more confounding factors. Thus, if clinically important interaction is considered likely, additional steps must be taken to permit the appropriate analyses (e.g., stratified randomization). This added complexity may result in a study that is no longer a truly simple trial.

When is a large simplified trial feasible?

LSTs are feasible when all of the conditions in Table 36.3 are met.

Table 36.3 Conditions which make a large, simple randomized trial 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

LSTs 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?"

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 compliance, 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, 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.

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.[29,30] Because of the participants' knowledge of medical conditions and symptoms and participation in the US health-care system, relatively sophisticated information could be obtained using mailed questionnaires, and even biologic samples could be collected. Success of the Boston University Fever Study 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.[24]

Logistics of conducting a large simplified trial

A 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 are "multicenter" studies involving a group of primary investigators who are responsible for the scientific conduct of the study, a central data coordinating facility, and a network of enrollment sites (possibly the offices of collaborating physicians or other health-care providers). Health-care professionals (e.g., physicians, nurse practitioners, and pharmacists in private practice or members of large health-care organizations) can participate by recruiting eligible patients. Alternative methods to identify and enroll eligible subjects (e.g., direct mailings to

professional groups, print ads) may be appropriate for some studies. Because success depends on the cooperation of multiple health-care 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.

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."[28] 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 and should ideally reflect only restrictions that would apply in a typical clinical setting.

Simple informed consent and registration documents should be completed, with one copy kept on file by the collaborating practitioner or study staff, one given to the study participant, and, depending on the role of the coordinating center, one forwarded to the data coordinating center by mail or facsimile. Registration of study subjects can also be accomplished online using a secure internet connection to the coordinating center, which allows for immediate confirmation of eligibility and randomization.[31] 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. Therefore, patients should be randomized only after eligibility has been confirmed and the enrollment process completed.

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.[29]

Importance of complete follow-up

Because drop outs 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, a study newsletter, or e-mail 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 follow-up data collection is the responsibility of the central study staff. Busy health-care providers cannot be expected to consistently obtain even minimal but specific 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) or current contact information for the occasional patient who would otherwise be lost to follow-up. A mailed questionnaire, supplemented by telephone interviews when needed, has been shown to work quite well.[24] 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. If there is a need to confirm a diagnosis or evaluate symptoms, a limited number of participants can be referred to their enrolling health-care provider for examination or to have studies performed. 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 randomization; any material differences between treatment groups suggest a failure of randomization. As noted above, it is usually assumed that there is no material interaction between patient characteristics and medication effects, thus eliminating the need for complex statistical analyses to test for effect modification.

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, it is appropriate 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 significant result cannot be achieved, even if the study were to be completed as planned. A data monitoring committee, independent of the study investigators, can conduct periodic reviews of the data using an appropriate group sequential analysis procedure to preserve the study's overall Type I error rate.[32,33]

The Boston University Fever Study, an office-based study of the safety of ibuprofen use in children, was an LST conducted by the authors using the approach described above. The methods and results have been described in detail but are briefly

summarized here.[1,24] The study was a practitioner-based, randomized clinical trial designed to compare the risk of rare but serious adverse events among children treated for fever with ibuprofen or acetaminophen. Study subjects 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). The primary study endpoints were hospitalization for gastrointestinal bleeding, acute renal failure, anaphylaxis, or Reye syndrome occurring within 4 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 gastrointestinal bleeds, all of whom had been assigned to receive ibuprofen—two in each dose group. The risk of hospitalization for gastrointestinal bleeding among all ibuprofen treated children, 7.2 per 100 000 (95% 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 over-the-counter sales of pediatric ibuprofen suspension in the United States.

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. This is particularly the case for drugs that are being considered for OTC switch, because the risks of rare and unknown events that would be acceptable under prescription status might be unacceptable when the drug is self-administered and likely used by diverse populations. 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 classical 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.[34]

One possible approach that may improve efficiency in large studies would be to conduct trials involving patients who receive care from very large health delivery systems with automated medical records. If reliable data concerning relevant outcomes (e.g., hospitalization for gastrointestinal bleeding) were available in automated medical records for all study participants, it would be theoretically possible to eliminate the need to contact patients to collect follow-up data. It would still be necessary to identify eligible subjects, obtain consent, and randomize treatment. In addition, assurance is needed that events were not being missed by patients presenting to out-of-plan providers.

In settings where there is no appropriate control treatment and it is not ethical to randomize between active drug and placebo, an alternative to an LST might be to enroll and follow a single cohort of

perhaps the first 10 000 users of a study medication. However, the absence of a comparison group would make it impossible to determine whether the observed risks were due to the drug, the disease, or other factors; however, although it would at least be possible to accurately estimate the absolute risk of important events, whatever their cause, among exposed subjects. An alternative and perhaps preferable approach would be to randomize to different doses, when possible, and search for a dose–response relationship.

It is clear that very large simple controlled trials of drug safety can be conducted.[2,24,25] It remains less clear, however, how frequently the factors that support the need for a very large trial (Table 36.2) will converge with those that permit such a trial to be carried out (Table 36.3). As a discipline, pharmacoepidemiology is well suited to conduct LSTs and to develop more efficient methods of subject 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. Mitchell AA, Lesko SM. When a randomized controlled trial is needed to assess drug safety: The case of pediatric ibuprofen. *Drug Saf* 1995; **13**: 15–24.
2. Hasford J, Bussmann W-D, Delius W, Koepcke W, Lehmann K, Weber E. First dose hypotension with enalapril and prazosin in congestive heart failure. *Int J Cardiol* 1991; **31**: 287–94.
3. Lesko S, Mitchell AA, Epstein MF, Louik C, Giacoia GP, Shapiro S. Heparin use as a risk factor for intraventricular hemorrhage in low-birth-weight infants. *N Engl J Med* 1986; **314**: 1156–60.
4. Malloy MH, Cutter GR. The association of heparin exposure with intraventricular hemorrhage among very low birth weight infants. *J Perinatol* 1995; **15**: 185–91.
5. Chang GY, Leuder FL, DiMichele DM, Radkowski MA, McWilliams LJ, Jansen RD. Heparin and the risk of intraventricular hemorrhage in premature infants. *J Pediatr* 1997; **131**: 362–6.
6. Rosenberg L, Armstrong B, Jick H. Myocardial infarction and estrogen therapy in post-menopausal women. *N Engl J Med* 1976; **294**: 1256–9.
7. Wilson PW, Garrison RJ, Castelli WP. Postmenopausal estrogen use, cigarette smoking, and cardiovascular morbidity in women over 50. The Framingham Study. *N Engl J Med* 1985; **313**: 1038–43.
8. Ross RK, Paganini-Hill A, Mack TM, Arthur M, Henderson BE. Menopausal oestrogen therapy and protection form death from ischaemic heart disease. *Lancet* 1981; **1**: 858–60.
9. Stampfer MJ, Willett WC, Colditz GA, Rosner B, Speizer FE, Hennekens CH. A prospective study of postmenopausal estrogen therapy and coronary heart disease. *N Engl J Med* 1985; **313**: 1044–9.
10. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, *et al.* 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.
11. Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. A policy perspective. *Ann Intern Med* 1979; **90**: 257–61.
12. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med* 1994; **331**: 394–8.
13. Francis T Jr, Korns R, Voight R, Boisen M, Hemphill F, Napier J, *et al.* An evaluation of the 1954 poliomyelitis vaccine trials: summary report. *Am J Public Health* 1955; **45** (Suppl): 1–50.
14. Gruppo Italiano per lo Studio della Streptochinasi nell'infarto miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986; **i**: 397–402.
15. Díaz R, Goyal A, Mehta SR, Afzal R, Xavier D, Pais P, *et al.* Glucose-insulin-potassium therapy in patients with ST-segment elevation myocardial infarction. *JAMA* 2007; **298**: 2399–405.
16. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, *et al.* Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009; **361**: 1139–51.
17. 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.
18. Fisher B, Costantion JP, Wickerham DL, Redmond CK, Kavannah M, Cronin WM, *et al.* Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 1998; **90**: 1371–88.

19. Heart Outcomes Prevention Evaluation (HOPE) Study Investigators. Vitamin E supplementation and cardiovascular events in high risk patients. *N Engl J Med* 2000; **342**: 154–60.
20. Klein EA, Lippman SM, Thompson IM, Goodman PJ, Albanes D, Taylor PR, *et al*. The selenim and vitamin E cancer prevention trial. *World J Urol* 2003; **21**: 21–7.
21. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* 1989; **321**: 129–35.
22. Thompson IM, Goodman PJ, Tangen CM, Lucia MS, Miller GJ, Ford LG, *et al*. The influence of finasteride on the development of prostate cancer. *N Engl J Med* 2003; **349**: 215–24.
23. ONTARGET Investigators, Yusuf S, Teo KK, Pogue J, Dyal L, Copland I, Schumacher H, *et al.* Telmisartan, ramipril, or both in patients at high risk for vascular events. *N Engl J Med* 2008; **358**: 1547–59.
24. Lesko SM, Mitchell AA. An assessment of the safety of pediatric ibuprofen: A practitioner-based randomized clinical trial. *JAMA* 1995; **273**: 929–33.
25. Van Ganse E, Jones JK, Moore N, Parc JM, Wall R, Schneid H. A large simple clinical trial prototype for assessment of OTC drug effects using patient-reported data. *Pharmacoepidemiol Drug Saf* 2005; **14**: 249–55.
26. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; **3**: 409–20.
27. Hasford J. Drug risk assessment: A case for large trials with lean protocols. *Pharmacoepidemiol Drug Saf* 1994; **3**: 321–7.
28. Gray R, Clarke M, Collins R, Peto R. Making randomized trials larger: a simple solution? *Eur J Surg Oncol* 1995; **2**: 137–9.
29. 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.
30. 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 Inst* 1999; **91**: 2102–6.
31. Santoro E, Nicolis E, Grazia Franzosi M. Telecommunications technology for the management of large scale clinical trials: the GISSI experience. *Comput Methods Programs Biomed* 1999; **60**: 215–23.
32. O'Brien PC. Data and safety monitoring. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics*. Chichester, UK: John Wiley and Sons, 1998, pp. 1058–66.
33. DeMets DL. Data and safety monitoring boards. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics*. Chichester, UK: John Wiley and Sons, 1998, pp. 1066–71.
34. Brass EP. The gap between clinical trials and clinical practice: the use of pragmatic clinical trials to inform regulatory decision making. *Clin Pharm Ther* 2010; **87**: 351–5.

CHAPTER 37
The Use of Pharmacoepidemiology to Study Beneficial Drug Effects

Brian L. Strom[1] *and Kenneth L. Melmon*[2]

[1] Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[2] Stanford University School of Medicine, Stanford, CA, USA

Introduction

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 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

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

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 anti-inflammatory action.[3] Only recently, relative to how long aspirin has been around, has this been confirmed as a valid new indication for the use of aspirin. A major research challenge is to discover this type of drug effect. For example, it currently remains an open question whether non-aspirin non-steroidal anti-inflammatory drugs have the same beneficial effects, although data are accumulating to that effect.[4] Secondarily, it is useful to quantitate the frequency of the event.

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 ingest 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 non-experimental study designs. However, these raise special methodologic problems, which are discussed below.

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, and is discussed in Chapter 38.

Note that the outcome variable for any of these studies can be of multiple different types. They can be clinical outcomes (diseased/ undiseased), or so-called "outcomes research," as defined by health services researchers (see Chapter 41 for a discussion of the validity issues involved in measuring such outcomes); they can be measures of quality-of-life (see Chapter 39), often referred to in the pharmaceutical industry as "outcomes research"; they can be measures of utility, that is global measures of the desirability of certain clinical outcomes (see Chapters 38 and 39); they can be economic outcomes (see Chapter 38); 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 (see Table 37.1).[5] Premarketing randomized clinical trials generally provide information on whether a drug can produce at least one beneficial effect. Specifically,

Table 37.1 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:
   a. Variations in drug regimen: dose per unit time, distribution of dose over time, duration of regimen
   b. Characteristics of the indication: severity, subcategories of the illness, changes over time
   c. 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 *et al.*[5] with permission from Elsevier.

premarketing studies generally investigate the efficacy of drug relative to a placebo, when 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 compliance (now more often called adherence) during these studies is assured, and the patients included are similar to each other in age and sex, do not have other diseases, and are not taking other drugs. Such restrictions 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 the drug can and does have other beneficial effects. In addition, at the time of marketing there may be little 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 the 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 a number of 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 37.2.[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

Table 37.2 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	6
Distribution of dose over time	Furosemide	Congestive heart failure	Efficacy improved by more frequent smaller doses	7
Duration	Clonidine	Hypertension	Tolerance develops in the absence of a diuretic	8
	Hypoglycemics (e.g., acetohexamide and tolazamide)	Diabetes mellitus	Tolerance develops in many patients	9
Indication				
Severity	Metaproterenol	Asthma	Patients with severe illness do not have a response without additional, supplementary therapy	10
Subcategories	Desipramine	Depression	May vary with endogenous versus exogenous depression	11
Changes over time	Ampicillin	Otitis media	No longer the drug of choice in some geographic areas due to bacterial resistance	12, 13
Patient				
Age	Diazepam	Anxiety	A given regimen is more effective in the aged than in the young	14
			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	15
Other illness	Gentamicin	Infection	Lower doses required in renal failure	16
Other				
Drugs	Lithium	Manic-depressive illness	Clearance impaired by diuretics, e.g., furosemide	17
	Acetohexamide	Diabetes mellitus	Many drugs interfere, by causing hyperglycemia (e.g., diuretics), displacing drug from binding sites (e.g., non-steroidal anti-inflammatory drugs), etc.	18
Diet	Diuretics (e.g., metolazone, furosemide)	Hypertension	A decrease in sodium intake can improve efficacy	19
	Lithium	Manic–depressive illness	Significant sodium depletion or excess can modify renal excretion	17

*Reproduced from Strom *et al.*[5] with permission from Elsevier.

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 per force are focused predominantly on meeting appropriate regulatory requirements, and only secondarily on providing a basis for optimal therapeutic decisions. The physician also should keep in mind that the FDA is not allowed to regulate physicians but, rather, pharmaceutical manufacturers. This regulation is not aimed at telling a physician precisely how an agent should be used. 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, that 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 47. 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 under study is the clinical condition that the drug is supposed to change (cure, ameliorate, or prevent). In clinical practice, 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.

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 post-myocardial 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, whether they are harmful or beneficial. 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 non-steroidal anti-inflammatory drugs, 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 non-steroidal anti-inflammatory drugs, the increased risk of hypersensitivity reactions evident in patients taking non-steroidal anti-inflammatory drugs 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 non-experimental 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 (see Table 37.3).[23] First,

Table 37.3 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 PPD
(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

Reproduced From Strom *et al.*[23] with permission from *Nature*.

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 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 will spontaneously 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.

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 are beginning to use *propensity scores* towards this end.[28,29] This is an approach that 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. When there are at least eight outcome events per confounder, logistic regression represents a preferable approach.[31] (See Chapter 47 for a more detailed discussion.)

Another 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; some would say impossible. This is discussed more in Chapter 47.

When questions of intended drug effects do not fall into any of the preceding categories, *confounding by the indication cannot be controlled*. Non-experimental 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. However, 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 37.3 and discussed further in Strom *et al*.[23]

Applicability of the proposed approaches

How commonly are the non-experimental 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 non-experimental 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, for example 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 non-experimental 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 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 non-experimental 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. Non-experimental 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 non-experimental 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 non-experimental 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 was impractical for many reasons. This is prophylactic therapy. 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 non-experimental studies were performed. Both case–control and cohort designs were used.[39–56] In general, these studies were rigorous and well done. Unfortunately, however, the question of confounding by the indication was not addressed in most of the studies.[33] In particular, most of the studies 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 routinely use the drugs and others might routinely avoid them.[63–65] In such a setting, non-experimental techniques could yield valid results, unaffected by confounding by the indication (Category 3.a.iv in Table 37.3). 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 are at high risk of complications from fractures. These situations would represent

uncontrollable confounding by the indication (Category 4 in Table 37.3). 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 trauma 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 non-experimental 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 massive clinical trial, that is 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 available 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 length, 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 6 weeks of outpatient treatment to 6 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 sub-

jects 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. There were a total of 3384 individuals 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 6 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 37.3). 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.

A decade later, a multicenter trial in Sweden, with 897 patients with a first episode of venous thromboembolism treated with oral anticoagulants and followed-up for 2 years, found a significant difference in the incidence of recurrent venous thromboembolism between the 6-week and 6-month groups (18.1% vs. 9.5%, respectively), and no significant difference in mortality or in 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

As another example, 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 confounding 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 37.3. Unfortunately, however, it is doubtful that 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 meta-analyses, indicating that lidocaine reduces ventricular fibrillation but increases mortality in acute myocardial infarction.[87] This was not confirmed in a subsequent paper, which re-analyzed 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 had been addressed using randomized clinical trials.[89] However, the results had been 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 37.3. 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 non-experimental study 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 to 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 that 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 on 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 compliance 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 non-experimental 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 is 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 is not being made by the prescriber on the basis of patient characteristics, but by the pharmacist on the basis of product availability—Category 3.a.ii in Table 37.3.

A few pharmacoepidemiologic studies (unpublished) on the relative effectiveness of different preparations used for the same purpose have been performed by Strom, using the COMPASS® database. These studies compared patients who were started on a brand name product and were switched to a generic product when it became available to 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 Section 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 later switch to a generic product. In each of the three studies, the future switchers were different from the future non-switchers, 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 studies 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 38.

Vaccines

Non-experimental 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] and 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 non-experimental study designs, the relatively indiscriminate use of such vaccines places the study in Category 2 of Table 37.3. Patients who receive these vaccines differ from those who do not in their socioeconomic status, their access to medical care, and their physicians' attitudes towards vaccines. However, for most vaccines, an individual physician is not likely to give only some of his 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. Non-experimental 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 non-experimental studies of vaccine efficacy.[148–154]

Cancer screening

Another frequent use of non-experimental study designs is to evaluate the efficacy of cancer screening programs. Although this does not directly relate to drugs, the methodologic implications are the same, and have been better enunciated than in the pharmacoepidemiologic literature. The use of non-experimental study designs to evaluate the efficacy of cancer screening programs will be briefly discussed here, therefore.

Once again, ideally, questions about the value of screening would be addressed using randomized clinical trials. However, most diseases that are screened for are relatively uncommon. Only a very small fraction of participants in a broad screening program could be expected to benefit from the screening program. Thus, randomized clinical trials of screening can be expensive and may require years to complete. Even more importantly, once a screening procedure is widely accepted, even without data documenting its efficacy, recruiting patients into a randomized clinical trial can be impractical and possibly truly unethical.

Instead, investigators have used non-experimental designs. Screening procedures that have been evaluated repeatedly in this fashion include the value of "Pap" smears for cervical cancer[155–170] and mammography and self-examination for breast cancer.[171–188] Other studies investigated screening measures for lung cancer,[189,190] gastric cancer,[191] prostate cancer,[192] ovarian cancer,[193] and colorectal cancer.[194] All of these were case–control studies. Again, they raise similar methodologic considerations of confounding by indication. Specifically, why do some women choose to have the screening procedure and others do not? One randomized clinical trial documented that women who attended screening sessions were at higher risk of developing breast cancer than women who were offered screening but did not attend.[195] In addition, case–control studies of screening present additional thorny methodologic problems regarding how to define cases, how to define controls, the time period to choose for the study, etc.[196–211]

Other examples

Other analogous work using case–control study designs has explored the effectiveness of bicycle safety helmets in preventing face injuries,[212,213] antibiotic prophylaxis in preventing postdental infective endocarditis,[214,215] beta-blockers in preventing mortality in patients with acute myocardial infarction,[216] beta-blockers and incident coronary artery events,[217] etc.

The future

Clinicians have long recognized the value of clinical observations and non-experimental research. Much of our current knowledge about the usefulness of medical interventions is based on information that is non-experimental. Yet the data and conclusions from the information are useful and valid. However, the information that observational techniques generate cannot be accepted uncritically. Perhaps in reaction to the limitations of non-experimental 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.* argue: "…to keep up with the clinical literature…discard at once all articles on therapy that are not randomized trials."[218] In light of the analysis presented above, this posture 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."[219] Many investigators are now applying non-experimental designs to studies of beneficial drug effects, and indeed an entire new field has emerged of comparative effectiveness research (see Chapter 32). 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 47), and hopefully more will be available in the future. However, when confounding by indication can be addressed, clinical observations and non-experimental research can be used. The results of non-experimental research are unlikely to be as powerful or as convincing as those of experimental research. We are not suggesting that non-experimental research, and certainly not suggesting that non-experimental 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 benefits, there frequently is a good alternative.

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, Jaskowiak J, Kishel L, Chittams J, *et al*. The effects of nonselective non-aspirin non-steroidal 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. Post-marketing 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, Reali E, Belvedere G, Standen SM, *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, PA: Lea & Febiger, 1979.
18. *Physicians' Desk Reference*, 33rd edn. Oradell, New Jersey: 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, Wahl P, Logerfo JP, Inui TS, *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 Int 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. *Springer Series in Statistics: Observational studies*. Springer-Verlag, New York, 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. Paul R, 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. Postmarketing studies of drug efficacy: how? *Am J Med* 1984; **77**: 703–8.
33. Strom BL. Are estrogen 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, Davies JL, Tonascia JA, Brown C, *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. Postmenopausal 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, Johnell O, Magnusson C, Persson PG, *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, Bunin G, Berlin JA, Baumgarten M, *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, Walker AH, Panossian S, Gallagher S, *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, Mouw T, Hollenbeck AR, Schatzkin A, *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* (Larchmt) 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, Cauley J, Ensrud K, Bauer DC, *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, Nielsen SP, Vestergaard P, Hermann AP, *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, Carlsson A, Lärfars G, Nicol 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, Rosenberg Y, Eby CS, Deitcher SR, *et al.* Long-term, low-intensity warfarin therapy for the prevention of recurrent venous thromboembolism. *NEJM* 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, Dyduszynski A, Kuch J, Nartowicz E, *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, Aylward PE, White HD, Thompson TD, *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, Baigent C, Remvig L, Mortensen S, *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. treat-

ments 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, Gordon S, Gangadharan V, Hauser AM, *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, Ochoa-Gondar O, de Diego C, Valdivieso A, *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, Torroba L, Navarro-Alonso JA, Irisarri F, *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, Weiss BP, Auerbach D, Habel LA, *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, Wahedi OK, Nahar N, Islam D, *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, Ziontz L, Christenson P, Waterman SH, *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, Hoppe CC, Makintubee S, Sitze SL, *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, Jacobs JL, Pierson LM, Neaton JD, *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**: 711–4.
127. Deming MS, Jaiteh KO, Otten MW Jr, Flagg EW, Jallow M, Cham M, *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, Ajello GW, Bolan GA, Adamsbaum C, *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, Evans D, Plikaytis BD, Wenger JD, *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, Strom BL, Wan CM, Liu HL, *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, Tsai TF, Wan CM, Tang SC, *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, Willett W, Fajardo I, Koch-Weser 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, Kumaraswamy RK, Mullick DN, Rungsung N, *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, Litkina IL, Filatov NN, Monisov AA, *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. 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, Hinman AR, Marks JS, Bart KJ, *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. Clarke EA, Anderson TW. Does screening by "PAP" smears help prevent cervical cancer? *Lancet* 1979; **2**: 1–4.
156. Aristizabal N, Cuello C, Correa P, Collazos T, Haenszel W. The impact of vaginal cytology on cervical cancer risks in Cali, Colombia. *Int J Cancer* 1984; **34**: 5–9.
157. LaVecchia C, Franceschi S, Decarli A, Fasoli M, Gentile A, Tognoni G. "Pap" smear and the risk of cervical neoplasia: quantitative estimates from a case-control study. *Lancet* 1984; **2**: 779–82.
158. MacGregor E, Moss SM, Parkin DM, Day NE. A case-control study of cervical cancer screening in North east Scotland. *BMJ* 1985; **290**: 1543–6.
159. van der Graaf Y, Zielhuis GA, Peer PG, Vooijs PG. The effectiveness of cervical screening: a population-based case-control study. *J Clin Epidemiol* 1988; **41**: 21–6.
160. Clarke EA, Hilditch S, Anderson TW. Optimal frequency of screening for cervical cancer: a Toronto case-control study. *IARC Sci Publ* 1986; **76**: 125–31.
161. Raymond L, Obradovic M, Riotton G. Additional results on relative protection of cervical cancer screening according to stage of tumour from the Geneva case-control study. *IARC Sci Publ* 1986; **76**: 107–10.
162. Geirsson G, Kristiansdottir R, Sigurdsson K, Moss S, Tulinius H. Cervical cancer screening in Iceland: a case-control study. *IARC Sci Publ* 1986; **76**: 37–41.
163. Berrino F, Gatta G, d'Alto M, Crosignani P, Riboli E. Efficacy of screening in preventing invasive cervical cancer: a case-control study in Milan, Italy. *IARC Sci Publ* 1986; **76**: 111–23.
164. Miller MG, Sung HY, Sawaya GF, Kearney KA, Kinney W. Hiatt RA. Screening interval and risk of invasive squamous cell cervical cancer. *Obstet Gynecol* 2003; **101**: 29–37.
165. Herrero R, Brinton LA, Reeves WC, Brenes MM, de Britton RC, Gaitan E, *et al.* Screening for cervical cancer in Latin America: a case-control study. *Int J Epidemiol* 1992; **21**: 1050–6.
166. Sato S, Makino H, Yajima A, Fukao A. Cervical cancer screening in Japan. A case-control study. *Acta Cytol* 1997; **41**: 1103–6.
167. Celentano DD, Klassen AC, Weisman CS, Rosenshein NB. Cervical cancer screening practices among older women: results from the Maryland cervical cancer case-control study. *J Clin Epidemiol* 1988; **41**: 531–41.
168. Celentano DD, Klassen AC, Weisman CS, Rosenshein NB. Duration of relative protection of screening for cervical cancer. *Prev Med* 1989; **18**: 411–22.
169. Zhang ZF, Parkin DM, Yu SZ, Esteve J, Yang XZ, Day NE. Cervical screening attendance and its

effectiveness in a rural population in China. *Canc Detect Prev* 1989; **13**: 337–42.
170. Sobue T, Suzuki T, Fujimoto I, Yokoi N, Naruke T. Population-based case-control study on cancer screening. *Environ Health Perspect* 1990; **87**: 57–62.
171. McCarthy EP, Burns RB, Freund KM, Ash AS, Shwartz M, Marwill SL, *et al.* Mammography use, breast cancer stage at diagnosis, and survival among older women. *J Am Geriatr Soc* 2000; **48**: 1226–33.
172. Wojcik BE, Spinks MK, Stein CR. Effects of screening mammography on the comparative survival rates of African American, white, and Hispanic beneficiaries of a comprehensive health care system. *Breast J* 2003; **9**: 175–83.
173. Collette JHA, Day NE, Rombach JJ, de Waard F. Evaluation of screening for breast cancer in a non-randomized study (the DOM project) by means of a case-control study. *Lancet* 1984; **1**: 1224–6.
174. Verbeek ALM, Hendriks JHCL, Holland R, Mravunac M, Sturmans F, Day NE. Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen project, 1975–1981. *Lancet* 1984; **1**: 1222–4.
175. Foster RS Jr, Costanza MC. Breast self-examination practices and breast cancer survival. *Cancer* 1984; **53**: 999–1005.
176. Greenwald P, Nasca PC, Lawrence CE, Horton J, McGarrah RP, Gabriele T, *et al*. Estimated effect of breast self-examination and routine physician examinations on breast cancer mortality. *N Engl J Med* 1978; **299**: 271–3.
177. Smith EM, Francis AM, Polissar L. The effect of breast self-exam practices and physical examination on extent of disease at diagnosis. *Prev Med* 1980; **9**: 409–17.
178. Huquely CM, Brown RL. The value of breast self-examination. *Cancer* 1981; **47**: 989–95.
179. Senie RT, Rosen PP, Lesser ML, Kinne DW. Breast self-examination and medical examination related to breast cancer stage. *Am J Public Health* 1981; **71**: 583–90.
180. Feldman JG, Carter AC, Nicastri AD, Hosat ST. Breast self-examination, relationship to stage of breast cancer at diagnosis. *Cancer* 1981; **47**: 2740–5.
181. Dubin N, Pasternack BS. Breast cancer screening data in case-control studies. *Am J Epidemiol* 1984; **120**: 8–16.
182. Dubin N, Friedman DR, Toniolo PG, Pasternack BS. Breast cancer detection centers and case-control studies of the efficacy of screening. *J Chronic Dis* 1987; **40**: 1041–50.
183. Collette HJA, Rombach JJ, Day NE, de Waard F. Evaluation of screening for breast cancer in a non-randomized study (the DOM project) by means of a case–control study. *Lancet* 1984; **1**: 1224–6.
184. Hunt KA, Rosen EL, Sickles EA. Outcome analysis for women undergoing annual versus biennial screening mammography: a review of 24,211 examinations. *Am J Roentgenol* 1999; **173**: 285–9.
185. Palli D, Del Turco MR, Buiatti E, Carli S, Ciatto S, Toscani L, *et al*. A case-control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). *Int J Cancer* 1986; **38**: 501–4.
186. Palli D, Del Turco MR, Buiatti E, Ciatto S, Crocetti E, Paci E. Time interval since last test in a breast cancer screening programme: a case-control study in Italy. *J Epidemiol Community Health* 1989; **43**: 241–8.
187. Friedman DR, Dubin N. Case-control evaluation of breast cancer screening efficacy. *Am J Epidemiol* 1991; **133**: 974–84.
188. Randolph WM, Goodwin JS, Mahnken JD, Freeman JL. Regular mammography use is associated with elimination of age-related disparities in size and stage of breast cancer at diagnosis. *Ann Intern Med* 2002; **137**: 783–90.
189. Ebeling K, Nischan P. Screening for lung cancer—results from a case-control study. *Int J Cancer* 1987; **40**: 141–4.
190. Sobue T, Suzuki T, Naruke T, and The Japanese Lung-Cancer-Screening Research Group. A case-control study for evaluating lung-cancer screening in Japan. *Int J Cancer* 1992; **50**: 230–7.
191. Oshima A, Hirata N, Ubukata T, Umeda K, Fujimoto I. Evaluation of a mass screening program for stomach cancer with a case-control study design. *Int J Cancer* 1986; **38**: 829–33.
192. Concato J, Wells CK, Horwitz RI, Penson D, Fincke G, Berlowitz DR, *et al*. The effectiveness of screening for prostate cancer: a nested case-control study. *Arch Intern Med* 2006; **166**: 38–43.
193. Saito M, Aoki D, Susumu N, Suzuki A, Suzuki N, Udagawa Y, *et al*. Efficient screening for ovarian cancers using a combination of tumor markers CA602 and CA546. *Int J Gynecol Cancer* 2005; **15**: 37–44.
194. Wahrendorf J, Robra BP, Wiebelt H, Oberhausen R, Weiland M, Dhom G. Effectiveness of colorectal cancer screening: results from a population-based case-control evaluation in Saarland, Germany. *Eur J Cancer Prev* 1993; **2**: 221–7.

195. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Selection, follow-up, and analysis in the Health Insurance Plan Study: A randomized trial with breast cancer screening. *Natl Cancer Inst Monogr* 1985; **67**: 65–74.
196. Connor RJ, Prorok PC, Weed DL. The case-control design and the assessment of the efficacy of cancer screening. *J Clin Epidemiol* 1991; **44**: 1215–21.
197. Morrison AS. Case definition in case-control studies of the efficacy of screening. *Am J Epidemiol* 1982; **115**: 6–8.
198. Weiss NS. Control definition in case-control studies of the efficacy of screening and diagnostic testing. *Am J Epidemiol* 1983; **118**: 457–60.
199. Sasco AJ, Day NE, Walter SD. Case-control studies for the evaluation of screening. *J Chronic Dis* 1986; **39**: 399–405.
200. Frommer DJ. Case-control studies of screening. *J Clin Epidemiol* 1988; **41**: 101.
201. Sasco AJ. Lead time and length bias in case-control studies for the evaluation of screening. *J Clin Epidemiol* 1988; **41**: 103–4.
202. Cole P, Morrison AS. Basic issues in population screening for cancer. *J Natl Cancer Inst* 1980; **64**: 1263–72.
203. Morrison AS. Case definition in case-control studies of the efficacy of screening. *Am J Epidemiol* 1982; **115**: 6–8.
204. Baum M, MacRae KD. Screening for breast cancer. *Lancet* 1984; **2**: 462.
205. Flanders WD, Longini IM Jr. Estimating benefits of screening from observational cohort studies. *Stat Med* 1990; **9**: 969–80.
206. Weiss NS, McKnight B, Stevens NG. Approaches to the analysis of case-control studies of the efficacy of screening for cancer. *Am J Epidemiol* 1992; **135**: 817–23.
207. Knox G. Case-control studies of screening procedures. *Public Health* 1991; **105**: 55–61.
208. Weiss NS. Case-control studies of the efficacy of screening tests designed to prevent the incidence of cancer. *Am J Epidemiol* 1999; **149**: 1–4.
209. Weiss NS. Analysis of case-control studies of the efficacy of screening for cancer: How should we deal with tests done in persons with symptoms? *Am J Epidemiol* 1998; **147**: 1099–102.
210. Cronin KA, Weed DL, Connor RJ, Prorok PC. Case-control studies of cancer screening: theory and practice. *J Natl Cancer Inst* 1998; **90**: 498–504.
211. Weiss NS, Lazovich D. Case-control studies of screening efficacy: the use of persons newly diagnosed with cancer who later sustain an unfavorable outcome. *Am J Epidemiol* 1996; **143**: 319–22.
212. 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.
213. 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.
214. Imperiale TF, Horwitz RI. Does prophylaxis prevent postdental infective endocarditis? A controlled evaluation of protective efficacy. *Am J Med* 1990; **88**: 131–6.
215. Strom BL, Abrutyn E, Berlin JA, Kinman JL, Feldman RS, Stolley PD, *et al*. Dental and cardiac risk factors for infective endocarditis. A population-based, case-control study. *Ann Intern Med* 1998; **129**: 761–9.
216. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; **89**: 630–8.
217. 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.
218. Sackett DL, Haynes RB, Tugwell PT. *Clinical Epidemiology. A Basic Science for Clinical Medicine*. Boston, MA: Little Brown and Co., 1985.
219. Hill AB. Reflections on the controlled trial. *Ann Rheum Dis* 1966; **25**: 107–13.

CHAPTER 38
Pharmacoeconomics: Economic Evaluation of Pharmaceuticals

Kevin A. Schulman,[1] Henry A. Glick,[2] Daniel Polsky,[2] and Shelby D. Reed[1]
[1] Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, USA
[2] Division of General Internal Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

Conventional evaluation of new medical technologies such as pharmaceutical products includes consideration of efficacy, effectiveness, and safety. Other chapters of this book describe in detail how such evaluations are carried out. The methodology is well developed, and drug regulation in developed countries requires studies of safety and efficacy to be performed prior to drug marketing (see Chapters 1 and 8). Health-care researchers from a variety of disciplines have also developed techniques for the evaluation of the economic effects of clinical care and new medical technologies. Clinicians, pharmacists, economists, epidemiologists, operations researchers, and others have contributed to the field of "clinical economics," an evolving discipline dedicated to the study of how different approaches to patient care and treatment influence the resources consumed in clinical medicine.[1–15]

The growth of clinical economics has proceeded rapidly as health policymakers have faced a continuing series of decisions about funding new clinical therapies in an era of increasingly constrained health-care resources. Assessments of new therapies include the resources required for the new therapy, the extent of the substitution of the new resources for existing resources, if any, and the health outcomes that result from therapeutic intervention. Thus, clinical economics includes not just an assessment of the cost of a new therapy, but a joint assessment of its overall economic and clinical effects.

This chapter discusses the need for applying economic concepts to the study of pharmaceuticals, introduces the concepts of clinical economics and the application of these concepts to pharmaceutical research, reviews some of the methodologic issues addressed by investigators studying the economics of pharmaceuticals, and finally offers examples of this type of research.

Clinical problems to be addressed by pharmacoepidemiologic research

There is ongoing concern about the cost of medical care, which has caused both purchasers and producers of pharmaceuticals to realize that the cost of drugs is not limited to their purchase price. The accompanying costs of preparation, administration, monitoring for and treating side effects, and the economic consequences of successful disease treatment are all influenced by the clinical and pharmacologic characteristics of pharmaceuticals. Thus, in addition to differences in efficacy and safety, differences in efficiency (or the effectiveness of the agent in actual clinical practice compared to its cost) distinguish drugs from one another.

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

Concerns about the cost of medical care in general and pharmaceuticals specifically are being felt in nearly all developed nations. A large number of national governments now require or are in the process of implementing requirements for the presentation of pharmacoeconomic data at the time of product registration for pharmaceuticals to qualify for reimbursement through the national health insurance systems.[16–26] Clinical economics research is being used increasingly by managed care organizations in the United States to inform funding decisions for new therapies.[27,28] At the local level, hospital administrators and other providers of health care are seeking ways of delivering high-quality care within the constraints of limited budgets or reduced fee schedules. These decision makers increasingly are interested in guidance regarding the cost-effectiveness of new medical technologies such as pharmaceuticals. This guidance can be provided by clinical economic analyses.

Trends in pharmacoeconomic research

The biotechnology revolution in medical research has added another challenge to pharmacoeconomic research. Pharmacoeconomics is increasingly being used to help determine the effect on patients of new classes of therapies before they are brought to the marketplace and to help determine appropriate clinical and economic outcomes for the clinical development program. The challenge is twofold: (i) understanding the potential effect of a therapy (e.g., whether a new antisepsis agent is a new type of antibiotic compound, where a short-term evaluation, efficacy at 14 days, is the appropriate clinical end point for analysis, or a life-supporting therapy, where a longer-term evaluation, efficacy at 6 or 12 months, is the appropriate clinical end point for efficacy assessment), and (ii) understanding the transition from efficacy to efficiency in clinical practice.[29,30] These challenges span the clinical development spectrum. As we learn more about the potential effects and use of a new product, these issues can be re-addressed in an iterative process. Finally, more and more firms are beginning to use economic models to help guide the business planning process and the new product development process to address the economic issues surrounding new therapies at the beginning of the product development cycle.

Pharmacoeconomic studies are designed to meet the different information needs of health-care purchasers and regulatory authorities. 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.[31]

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.

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. New pharmaceuticals are developed in a series of well-defined stages due to the regulatory process of drug approval (see Chapters 1 and 8). After a compound is identified and thought to be clinically useful, four distinct sets of evaluations—referred to as Phase I through IV studies—are mandated by the US Food and Drug Administration (FDA) and most other equivalent regulatory bodies. Phase I studies represent the first introduction of a new compound into (usually undiseased) humans, principally for the evaluation of safety and dosage. In Phase II studies, the drug is introduced into a patient population with the

disease of interest, again principally for the evaluation of safety and dosing. Phase III studies are randomized trials evaluating the safety and efficacy of new drugs, compared either with placebo or with a therapy that the new drug might replace (in the US, the appropriate comparator often is the subject of negotiations between the developer of the drug and the FDA). In addition to these three types of studies, drugs often are evaluated after they are marketed in what are referred to as Phase IV or postmarketing studies.

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.[32] Phase III studies can include economic assessments of new therapies as a primary or secondary end point (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 changes in hospitalization rates).[33–39]

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 Chapters 32 and 36).[40] These postmarketing studies may include assessments of the new therapy compared with "usual care" or compared with specific therapeutic agents. Again, the economic analysis can serve as a primary or secondary end point of the study.

Developing economic data as an end point 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 end points 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 end points for study (e.g., direct, productivity, and intangible costs to patients and caregivers, as well as 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 the 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.

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 review several studies that illustrate the usefulness of pharmacoeconomic research.

Methodologic problems to be addressed by pharmacoepidemiologic research

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 forgone. 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 forgone (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 38.1 with which readers should become familiar.[1] Along the X-axis are three types of economic analysis—cost identification, 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 equal, an investment should be undertaken when its benefits exceed its costs.

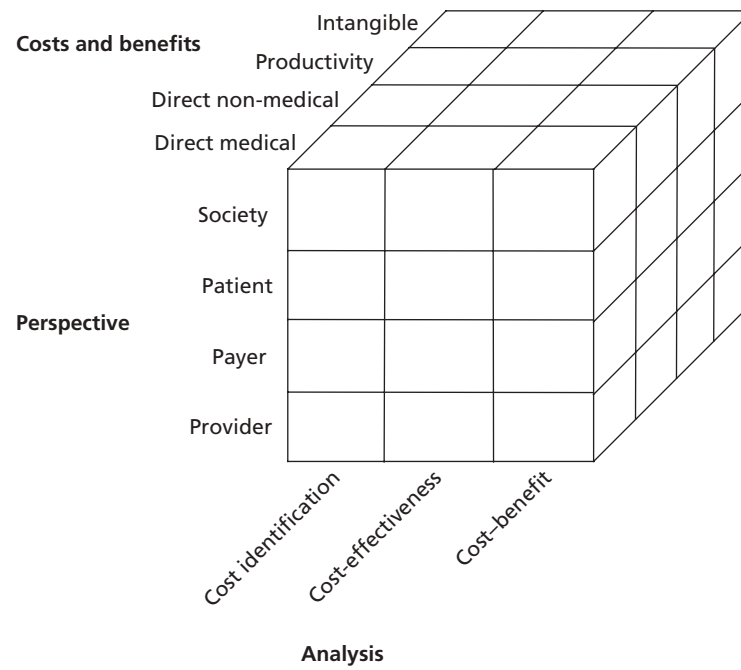


Figure 38.1 The three dimensions of economic evaluation of clinical care. Reproduced from Bombardier *et al.*[1] with permission from Journal of Rheumatology.

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.

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 health-care analyses. Outcomes (treatment benefits) may be difficult to measure in units of currency. 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 quality-adjusted life-years (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 dominant and should always be adopted. Programs that cost more and are more effective should be adopted if their incremental cost-effectiveness ratios fall within an acceptable range[41–43] and the budget for the program is acceptable. 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.

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 generally is 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, 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).[41–43] 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 performing cost identification analysis, 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 38.1 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 that are 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.[7,44–48]

Direct non-medical costs

Monetary transactions undertaken as a result of illness or health care to detect, prevent, or treat disease are not limited to direct medical costs. There is another type of cost that often is overlooked: direct non-medical 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., the cost of modification of a home to accommodate an ill individual). Direct non-medical costs, which are generally paid out of pocket by patients and their families, are just as much direct medical costs as are expenses that are more usually covered by third-party insurance plans.

Productivity costs

In contrast to direct costs, productivity 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 work force). 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 health care.[49–53] Sometimes, as with varicella vaccination,[54] the productivity costs of an illness are substantially greater than the direct costs of the illness.[55,56]

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[3]—at present these costs are often omitted in clinical economics research.

Perspective of analysis

The third axis in Figure 38.1 is that of 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.[15] The cost to society is the opportunity cost, the value of the opportunities forgone because of the resource having been 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 may have 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.[57,58] 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.

Methodologic issues in the pharmacoeconomic assessment of therapies

The basic approach for performing economic assessments of pharmaceutical products, as discussed above, has been adapted from the general methodology for cost-effectiveness and cost–benefit analysis. These methods have been well developed in medical technology assessment as well as in other fields of economic research. However, there remain a number of methodologic issues that confront investigators in economic evaluations of pharmaceutical therapies. This section reviews some of these issues as they arise in the design, analysis, and interpretation of pharmacoeconomic evaluations.

Clinical trials versus common practice

One of the most vexing of these issues is how to assess the cost implications of products during clinical trials. Ascertaining whether or not a product's costs are adequately offset by its effects or benefits presents a number of issues for consideration.[59]

The problem

Clinical trials are useful for determining the efficacy of therapeutic agents (see Chapters 3 and 36). However, their focus on efficacy rather than effectiveness (see Chapters 32 and 37) and their use of protocols for testing and treating patients pose problems for cost-effectiveness analysis. One difficulty in assessing the economic impact of a drug on an end point in a clinical trial is the performance of routine testing to determine the presence or absence of a study outcome. For example, in a study of prophylaxis against thromboembolic events, the protocol may specify testing of all patients for deep vein thromboses (DVT) (e.g., fibrinogen scanning, venograms, or Doppler testing), whether or not the patients show clinical signs of these events. While this diagnostic strategy may be appropriate, it is not necessarily common practice. Yet, it can have wide-ranging effects on the calculated costs and outcomes of care.

First, the protocol may induce the detection of extra cases—cases that would have gone undetected if no protocol were used in the usual care of patients. These cases may be detected earlier than they would have been in usual care. In the prophylaxis example above, repeated testing of all patients is likely to increase the number of DVTs that are detected, especially if, in usual care, patients are only tested when they develop clinical symptoms or signs of DVT. This extra or early detection may also reduce the average costs for each case detected, because subclinical cases or those detected early may be less costly to treat than clinically detected cases. However, because these two potential biases—more cases, each of which may cost less—work in opposite directions, the total costs of care for the patients in the trial may or may not exceed those that would occur in usual care.

Second, protocol-induced testing may lead to the detection of adverse drug effects that would otherwise have gone undetected. As above, the average costs of each may be less because the adverse effects would be milder. However, their frequency would obviously be higher, and they could result in additional testing and treatment.

Third, protocol-induced testing also may lead to the occurrence of fewer adverse events from the pharmaceutical product than would occur in usual care. The extra tests done in compliance with the protocol may provide information that otherwise would not have been available to clinicians, allowing them to take steps to prevent adverse events and their resulting costs. For example, an antibiotic protocol may call for more frequent testing of creatinine levels than would be conducted in usual care. These tests may warn physicians of impending renal problems, allowing them to change the drug dosage or the antibiotic. Thus, cases of nephrotoxicity that would have occurred in usual care may be avoided. This potential bias of reducing the costs of side effects and adverse events would tend to lower the overall costs of care observed in the trial compared to usual care.

Fourth, due to ethical obligations that arise when patients are enrolled in trials, outcomes detected in trials may be treated more aggressively than they would be in usual care. In trials, it is likely that physicians will treat all detected treatable clinical outcomes. In usual care, physicians may treat only those outcomes that in their judgment are clinically relevant. This potential bias would tend to increase the costs of care observed in the trial compared to usual care.

Fifth, protocol-induced testing to determine the efficacy of a product or to monitor the occurrence of all side effects, whether clinically detectable or not, generally will increase the costs of diagnostic testing in the trial, because many of these tests likely would be omitted in usual care. Alternatively, the protocol may reduce these costs in environments where there is overuse of testing. In teaching settings, for example, some residents may normally order more tests than are needed, and this excess testing may be limited by the protocol's testing prescriptions.

Sixth, clinical protocols may offer patients additional resources that are not routinely available in clinical practice. These additional resources may provide health benefits to patients. For example, protocols offering extensive home care services may affect the observed benefits of a therapy if the nursing intervention improves the management of the patient's illness. This could result in a bias in the study design if there are differences in the amount of home care services provided to patients in the treatment and control arms of a trial, or may result in additional health benefits to all study patients.

Seventh, patients in trials often are carefully selected. If a study sample has a mean patient age of 45 years, the result of the trial may not be readily generalizable to substantially older or younger populations. Similarly, exclusion criteria in clinical protocols may rule out patients with specific clinical syndromes (e.g., diabetes mellitus), women of childbearing potential, or patients of advanced age. These patients may require additional resources or may receive less benefit from therapy because their remaining life expectancy is shorter than that of patients enrolled in clinical trials. These exclusions further limit the generalizability of the findings of efficacy studies.

A related issue in pharmacoeconomics trials is the generalizability of the health care delivery system of the patients in the study. A pharmacoeconomic study conducted through an HMO using its members as subjects may observe fewer referrals to specialist physicians than would the same clinical study in a different practice setting. This effect may be even more pronounced in multinational clinical trials, where health care systems, physician education, and patients' expectations for treatment differ by country.

Other difficulties in projecting the results of clinical trials to usual care arise because the patients in clinical trials generally comply more completely with their treatment than do patients in usual care, because they receive prescribed patterns of care, and because trials often have a placebo arm. If there is an actual placebo effect, this last factor may tend to understate the effectiveness the agent will have when it is utilized in usual care.

Routinely appending economic evaluations to clinical trials will likely yield "cost-efficacy" analyses, the results of which may be substantially different from the result of cost-effectiveness analyses conducted in the usual care setting. The problem of generalizability is similar to that found in clinical epidemiologic research. Clinical economics explic-

itly recognizes the added complexity of having different resource-induced costs and benefits derived from clinical protocols and from observing patients in different health care systems in multicenter clinical trials.[60]

Pharmacogenomic strategies offer opportunities to segment populations of patients according to clinical benefit or risk[61] (see Chapter 34). In addition to evaluating the cost-effectiveness of individual therapies in select patient groups, studies should evaluate the incremental costs and consequent health outcomes resulting from the incorporation of pharmacogenomic information into treatment strategies.[62] Application of these new tools to the clinical development program will allow for tailoring of therapies in ways that could alter economic evaluations of therapies.[30]

Possible solutions

One possible solution to the problem of determining the economic impact of a drug in a clinical trial will be illustrated by examining the impact of a "usual care" arm appended as a third arm of a clinical trial. In such a three-arm study, patients randomized to the usual care arm of the study would be treated as they would be outside of the trial, rather than as mandated by the study protocol, and economic and outcomes data from usual care could thus be collected. These data would make it possible to quantify the number of outcomes that likely would be detected in usual care and the costs of these outcomes.

One drawback to this method is that physicians in the trial may treat all patients similarly, whether they are in the protocol-driven arm or the usual care arm of the study. This contamination can be partially overcome by randomizing physicians to the protocol or usual care arms, and can be overcome more completely by randomizing the sites of care (e.g., different hospitals for different arms of the study). However, these options require large numbers of physicians and/or sites of care and, thus, are very costly to implement. Moreover, such a strategy may result in non-random assignment of patients to treatment arms.

A second method that has been used to overcome these problems[7] is to collect data from patients who are not in the trial but who would have met its entry criteria, using these data to estimate the likely costs and outcomes in usual care. These patients could have received their care prior to the trial (historical comparison group) or concurrent with it (concurrent comparison group). In either case, some of the data available in the trial may not be available for patients in the comparison groups. Thus, investigators must insure comparability between the data for usual care patients and trial patients.

Two problems arise when using a concurrent comparison group to project the results of a trial to usual care. First, as with the randomization scheme above, the use of a protocol in the trial may affect the care delivered to patients who are not in the trial. If so, usual care patients may not receive the same care they would have received if the trial had not been performed. Thus, the results of the trial may lose generalizability to other settings. Second, the trial may enroll a particular type of patient (e.g., investigators may "cream-skim" by enrolling the healthiest patients with the fewest complications), possibly leaving a biased sample (e.g., of sicker and more complicated patients) for inclusion in the concurrent comparison group. This potential bias would tend to affect the estimate of the treatment costs that would be experienced in usual care.

Adoption of a historical comparison group would offset the issue of contamination. Because the trial was not ongoing when these patients received their care, it could not affect how they were treated. A historical comparison group would also tend to offset the selection bias: the subset of patients who would have been included in the trial if it had been carried out in the historic period will be candidates for the comparison group. However, use of a historic comparison group is unlikely to offset this bias entirely. Because this group is identified retrospectively, its attributes likely will reflect those of the average patients eligible for the trial, rather than those of the subset of patients that would have been enrolled in the trial (e.g., if cream-skimming had occurred).

In addition, differences between the care provided to patients in the trial and that provided to

patients in the historical group may be due as much to secular trends in the provision of medical care as they are to the adoption of a study protocol. For example, hospital length of stay in the United States has decreased since the early 1980s, due in part to the implementation of the Medicare Prospective Payment System. Thus, historical cohorts from earlier periods may have had longer lengths of stay as inpatients than is currently seen in clinical practice. These data may suggest a protocol-induced decrease in length of stay when one really does not exist.

To avoid these difficulties, the usual care comparison group may include both historic and concurrent comparison groups. In this case, multivariable methods such as multiple regression analysis or other analytic techniques must be used to control as best as possible for differences among the historic and concurrent comparison groups as well as between the comparison groups and the patients in the trial. For example, in a regression analysis of length of stay in the trial and in usual care, variables representing each of the groups will indicate the magnitude of the secular trends, the selection bias, and the protocol effects of the trial.

A number of methods currently are being investigated to help overcome the potential biases of resource-induced costs and benefits in clinical trials. These approaches include the development of "large and simple clinical trials" or pragmatic trials (see Chapter 36) that attempt to study real-world patient populations and conducting the trial in different health systems simultaneously to assess the impact of the therapy in different delivery settings (e.g., using a large HMO as a clinical testing site). Also, as opposed to randomized clinical trials, which often include a placebo comparator for the purpose of obtaining regulatory approval, pragmatic trials include clinically relevant comparator arms and focus on health outcomes as opposed to more limited end points.[63]

Issues in the design of prospective pharmacoeconomic studies

We have already addressed some of the general issues in the design and interpretation of pharmacoeconomic studies. Yet, prospective pharmacoeconomic studies, especially within Phase III clinical trials, are often our only opportunity to collect and analyze information on new therapeutic products before decisions are made concerning insurance reimbursement and formulary inclusion for these agents. We now address issues that arise in the design of these studies.

Sample size

The formula for estimating the sample size needed to test whether the incremental cost-effectiveness ratio falls within an acceptable range is as follows:[64,65]

$$n = \frac{2(z_\alpha + z_\beta)^2\left(sd_c^2 + (Wsd_q)^2 - (2W\rho sd_c sd_q)\right)}{(WQ - C)^2}$$

where n is the sample size per treatment group; z_α and z_β are the z-statistics for the α (e.g., 1.96 for two-tailed 95% confidence) and β (e.g., 0.84 for one-tailed 80% power) errors; *sd* is the standard deviation for cost (sd_c) and effect (sd_q); W equals the maximum willingness to pay for the outcome (i.e., the upper bound on what is considered reasonable), and ρ is the correlation of the difference in cost (C) and effect (Q). This formula identifies the sample size required to test a hypothesis such as that the point estimate of the cost-effectiveness ratio will be less than $75 000 per QALY (W).

Often those setting up clinical trials focus on the primary clinical question when developing sample-size estimates (see Chapter 4). They fail to consider the fact that the sample required to address the economic questions posed in the trial may differ from that needed for the primary clinical question. The formula for calculating the power to test the hypothesis that the resulting ratio will be acceptable is as follows:

$$z_\beta = \sqrt{\frac{n(WQ - C)^2}{2\left(sd_c^2 + (Wsd_q)^2 - (2W\rho sd_c sd_q)\right)}} - z_\alpha$$

In some cases, the sample size required for the economic analysis is smaller than that required to address the clinical question. More often, however, the opposite is true, in that the variances in cost and patient preference data are larger than those

for clinical data. Then one needs to confront the question of whether it is either ethical or practical to prolong the study for longer than need be to establish the drug's clinical effects.

Participation of patients

Those planning Phase III clinical trials usually are more focused on the clinical results of the trial than they are on the economic results; they would usually like to keep the number of centers and subjects needed to complete the trial to a minimum; and they would rather finish the trial sooner than later. Thus, they have a concern that patients might agree to participate in the clinical trial, but not be willing to participate in the economic portion of the trial. In such a case, the investigators often argue that patients should be allowed to participate in the clinical portion of the trial but be excluded from the economic portion of the trial. While self-selection always poses difficulties for trials, it should be clear that this suggestion is particularly worrisome. The economic assessment would end up comparing an estimate of effects from the entire sample with an estimate of costs from a non-random subset of the entire sample, thus allowing substantial bias to enter the analysis. Protocols should allow prospective collection of resource consumption and patient preference data, while sometimes incorporating a second consent to allow access to patients' financial information. This second consent would be important if the primary concern was that the patients included in the economic evaluation would not be representative of the entire cohort of patients participating in the trial. However, given the low rates of refusal to the release of financial information, a single consent form should be preferred for all trial data. The single consent avoids the possibility of selection bias in the economic end points relative to the clinical end points.

Data collection

As incorporation of an economic evaluation becomes more commonplace in clinical trials, those involved in the pharmacoeconomic evaluation are asked to incorporate their data collection needs into the trial protocol and the trial case report form. Collection of resource consumption data from primary or secondary sources is essential for a prospective economic evaluation of a pharmaceutical therapy. Some data elements, such as patient preference assessments, can only be collected on a prospective basis. Other data elements, such as medical resource use for patients in a particular health system could be ascertained retrospectively. In any case, data needs for the eventual study require comprehensive evaluation at the outset.

While some prospective data collection is required for almost all pharmacoeconomic studies, the amount of data to be collected for the pharmacoeconomic evaluation is still the subject of much debate.[32] There is no definitive means of addressing this issue at present. Phase II studies can be used to develop data that will help determine which resource consumption items are essential for the economic evaluation. Without this opportunity for prior data collection, however, we have to rely upon expert opinion to suggest major resource consumption items that should be monitored within the study. Duplicate data collection strategies (prospective evaluation of resource consumption within the study's case report form with retrospective assessment of resource consumption from hospital bills) can be used to ensure that data collection strategies do not miss critical data elements. However, one should develop a plan *a priori* to address any potential inconsistencies between information coming from different sources.

Resources are divided into specific categories for assessment for prospective data collection: inpatient resource use, outpatient resource use, and non-acute-care resource use. Within each of these categories, data can be subdivided into several categories: professional services (physicians, nurses, allied health professionals), hospital setting (intensive care unit, step-down unit, general medical floor), major diagnostic tests (radiologic tests, laboratory tests, nuclear medicine studies), major surgical procedures (operations and non-operating room procedures), and medications. Sample data collection forms for inpatient and outpatient resource consumption are presented as Figures 38.2 and 38.3. Issues related to data

HOSPITAL DISCHARGE FORM		Patient No. __ __ __ __
Principal Investigator _____	Study Hospital _____	Date of Admission _______ Date of Discharge _______
Source of Admission [ ] Emergency room [ ] Transfer (from _____) [ ] Elective	Discharge Diagnosis 1._____ 2._____	

Unit Type	Number of Days
Intensive Care Unit	
Intensive Care Unit with Mechanical Ventilator	
Bone Marrow Transplant Unit	
Step-Down/Intermediate Care Unit	
General Medical or Surgical Floor	

Pharmacologic Therapy	Total Dose
[ ] Study Drug _____	
[ ] Control _____	
Continuous IV Medication _____	

Types of Procedures	Date

Diagnostic Tests	Number of Tests
MRI	
CT Scan	
Bone Scan	

Figure 38.2 Inpatient resource assessment. This is a sample case report form for prospective assessment of inpatient resource consumption in a pharmacoeconomic study.

OUTPATIENT VISIT RECORD	**Patient No.** ___ ___ ___ ___

Name of Physician and Location of Visit (e.g., Emergency Room, Outpatient Clinic, Day Surgery, Home, Office)	Duration (in minutes)				
	date ____	date ____	date ____	date ____	date ____
1.					
2.					
3.					
Name of Nurse Clinician and Location of Visit (e.g., Emergency Room, Outpatient Clinic, Day Surgery, Home, Office)	date ____	date ____	date ____	date ____	date ____
1.					
2.					

Type of Procedure	Date
1.	
2.	
3.	

Diagnostic Tests	Number of Tests
MRI	
CT Scan	
Bone Scan	
Other	

Other Therapy (medications, etc.)	Date
1.	
2.	
3.	

Figure 38.3 Outpatient resource assessment. This is a sample case report form for prospective assessment of outpatient resource consumption in a pharmacoeconomic study.

collection for economic studies have been reviewed elsewhere.[66]

Appropriate comparators

Selection of appropriate treatment alternatives in a clinical study is essential for a useful economic evaluation of a pharmaceutical therapy. This issue is both a clinical and an economic one. Comparators can be the most common alternative therapies for a condition, or the lowest possible cost alternatives, even when not frequently used. However, in pharmacoeconomic studies, higher-cost treatment and/or less effective comparators may be inappropriately selected such that incremental costs between treatment alternatives would be smaller and incremental effectiveness between treatments would be larger, resulting in a lower cost-effectiveness ratio. Phase III studies have special limitations in this regard, because agents will typically be compared against the placebo to assess efficacy rather than against alternative treatments to assess the relative effectiveness of the agent.

Multicenter evaluations

The primary result of economic evaluations usually is a comparison of average, or pooled, differences in costs and differences in effects among patients who received the therapies under study. It is an open question, however, whether pooled results are representative of the results that would be observed in the individual centers or countries that participated in the study.[38,60,67] In some, the therapy may provide good value for the cost, while in others it may provide poor value. Three reasons commonly cited for these differences are differences in practice patterns (i.e., medical service use), differences in absolute and relative prices for medical service use (i.e., unit costs), and differences in underlying morbidity/mortality patterns in different centers and countries.[68–71]

There is a growing literature that addresses the transferability of a study's pooled results to subgroups.[38,60,68,70,72] Approaches include evaluation of the homogeneity of different centers' and countries' results,[73,74] use of random effects models to borrow information from the pooled results when deriving center-specific or country-specific estimates, direct statistical inference by use of net monetary benefit regression,[75–77] and use of decision analysis.[78]

Factors affecting resource consumption

Pharmacoeconomic research holds as a basic assumption the proposition that clinical severity of disease is the sole determinant of resource use by patients. Studies of regional variation highlight the shortcomings of this assumption.[79–81] This creates a significant challenge for health services research, and for pharmacoeconomics in particular. For example, when a new therapy is introduced to reduce severity of disease as a substitute for physician services that similarly reduce the severity of disease, if physicians either continue to provide the service to maintain their clinical practice or change the characteristics of the patients to whom they provide services (e.g., perform surgery on less severely ill patients), we will not achieve the full potential economic advantage afforded by the new therapy.

Economic data

Analysts generally have access to resource utilization data such as length of stay, monitoring tests performed, and pharmaceutical agents received. When evaluating a therapy from a perspective that requires cost data rather than charge data, however, it may be difficult to translate these resources into costs. For example, does a technology that frees up nursing time reduce costs, or are nursing costs fixed in the sense that the technology is likely to have little or no effect on the hospital payroll? Economists taking the social perspective would argue that real resource consumption has decreased and thus nursing is a variable cost. Accountants or others taking the hospital perspective might argue that, unless the change affects overall staffing or the need for overtime, it is not a saving. This issue depends in part on the temporal perspective taken by the analyst. In the short term, it is unlikely that nursing savings are recouped; in the long term, however, there probably will be a redirection of services. This analysis may also be confounded by the potential increase in the quality of care that nurses with more time may be able to

provide to their patients. In countries that have a shortage of hospital beds, hospital administrators often do not recognize staffing savings from early discharge programs, because the bed will be occupied by a new patient as soon as the old patient is discharged.

Perspective

When perspectives other than the societal perspective are adopted, it is unclear which benefits or outcomes should be counted in the analysis. For example, if a governmental agency's perspective is adopted, in which transfer payments such as pensions are counted as costs, quick deaths at age 65 may be valued more than long, costly deaths at age 75. Independent of whether we should condone the practice of using studies that are limited to costs in this type of scenario in decision making, it should be recognized that a cost-effectiveness analysis would account for health benefits gained by avoiding death at age 65.

In summary, due to their focus on efficacy and their use of clinical protocols, economic assessments of pharmaceutical products based upon Phase III clinical trials are not without their problems. However, these issues can be developed in pharmacoeconomic analysis plans and addressed prospectively or through supplemental data collection activities conducted concurrently with the clinical trial.

Measurement and modeling in clinical trials

Previously, we have discussed the development of pharmacoeconomic data throughout the drug development process. However, the types of data available at the end of the trial will depend upon the trial's sample size, duration, and clinical end points.

There are two categories of clinical end points considered in pharmacoeconomic analysis: intermediate end points and final end points. An intermediate end point is a clinical parameter, such as systolic blood pressure, which varies as a result of therapy. A final end point is an outcome variable, such as change in survival or quality-adjusted survival, which is common to several economic trials, which allows for comparisons of economic data across clinical studies, and is of relevance to policy makers.

The use of intermediate end points to demonstrate clinical efficacy is common in clinical trials, because it reduces both the cost of the clinical development process and the time needed to demonstrate the efficacy of the therapy. Intermediate end points are most appropriate in clinical research if they have been shown to be related to the clinical outcome of interest, as in the following:

- the use of changes in blood cholesterol levels to demonstrate the efficacy of new lipid lowering agents (intermediate end point: changes in low-density and high-density lipoprotein levels; final end point: changes in myocardial infarction rate and survival; demonstration of the relationship between intermediate and final end points: Framingham Heart Study[82]); and
- the use of change in blood pressure to demonstrate the efficacy of new antihypertensive agents (intermediate end point: changes in systolic and diastolic blood pressure; final end point: changes in stroke rates and survival; demonstration of the relationship between intermediate and final end points: Framingham Heart Study[83]).

Ideally, a clinical trial would be designed to follow patients throughout their lives, assessing both clinical and economic variables, to allow an incremental assessment of the full impact of the therapy on patients over their lifetimes. Of course, this type of study is almost never performed. Instead, most clinical trials assess patients over a relatively short period of time. Thus, most pharmacoeconomic assessments must utilize data collected from within the clinical trial in combination with an epidemiologic model to project the clinical and economic trial results over an appropriate period of a patient's lifetime.

The importance of this effort is illustrated in the following hypothetical example. A new therapy is under development that reduces the absolute risk of dying from a chronic disease by 50% as measured in a 1-year trial. However, this therapy is not curative. A 4-year trial was initiated at the same time as the 1-year trial. The first-year results were the same in both the 4-year trial and the 1-year trial. However, there was an increased risk of death

for treatment patients in the second and third year of the 4-year trial, and by the end of the third year of the trial the survival rate was identical in the treatment and control arms of the 4-year trial. While there was a clear benefit to the new therapy in terms of postponing events from the first year of treatment to later years, the economic assessment of the therapy would suggest a greatly reduced treatment benefit from the 4-year trial as compared with the 1-year trial.

In projecting results of short-term trials over patients' lifetimes, it is typical to present at least two of the many potential projections of lifetime treatment benefit.[84] A one-time effect model assumes that the clinical benefit observed in the trial is the only clinical benefit received by patients and does not persist after the trial. Under this model, after the trial has ended, the conditional probability of disease progression for patients is the same in both arms of the trial. Given that it is unlikely that a therapy will lose all benefits as soon as one stops measuring them, this projection method generally is pessimistic compared to the actual outcome. A continuous-benefit effect model assumes that the clinical benefit observed in the trial is continued throughout the patients' lifetimes. Under this model, the conditional probability of disease progression for treatment and control patients continues at the same rate as that measured in the clinical trial. In contrast to the one-time model, this projection of treatment benefit most likely is optimistic compared to the true long-term treatment outcome.

While we and others have developed models as secondary analyses of new therapies,[29,30,84–87] it is now common for clinical trials to incorporate an economic evaluation that involves primary data collection.[32,34–37,88–95] In fact, reviews of findings from economic evaluations are increasingly reported in the literature.[96–98] This change has resulted from an increasing awareness of the need for reliable economic data about new therapies at the time when the therapies are being introduced to the market. This impetus has also resulted from issues related to the complexity and cost of developing appropriate economic data for a secondary analysis of a new therapy, and issues related to the potential for bias in the design of economic studies conducted from analysis of secondary data sources.[16,17,99–101] However, as illustrated above, even primary data collection in clinical trials does not eliminate the need for treatment models in the economic analysis of new therapies.

Analysis plan for cost data

Analysis of cost data shares many features with analysis of clinical data. One of the most important is the need to develop an analysis plan before performing the analysis. Table 38.1 identifies a set of tasks that should be addressed in such a plan. The analysis plan should describe the study design (e.g., report on whether the trial is randomized and double-blind; identify the randomization groups; outline the recruitment strategy; describe the criteria for patient evaluation), and any implications the design has for the analysis of costs (e.g., how one will account for recruiting strategies such as rolling admission and a fixed stopping date).

The analysis plan should also specify the hypothesis and objectives of the study, define the primary and secondary end points, and describe how the end points will be constructed (e.g., multiplying resource counts measured in the trial times a set of unit costs measured outside the trial). In addition, the analysis plan should identify the potential covariables that will be used in the analysis and specify the time periods of interest (e.g., costs and clinical outcomes at 6 months might be the primary

Table 38.1 Steps in an economic analysis plan

(1)	Study design/summary
(2)	Study hypothesis/ objectives
(3)	Definition of endpoints
(4)	Covariates
(5)	Prespecification of time periods of interest
(6)	Statistical methods
(7)	Types of analyses
(8)	Hypothesis tests
(9)	Interim analyses
(10)	Multiple testing issues
(11)	Subgroup analyses
(12)	Power/ sample size calculations

outcome, while costs and clinical outcomes at 12 months might be a secondary outcome).

Also, the analysis plan should identify the statistical methods that will be used and how hypotheses will be tested (e.g., a *p*-value cutoff or a confidence interval for the difference that excludes 0). Further, the plan should prespecify whether interim analyses are planned, indicate how issues of multiple testing will be addressed, and predefine any subgroup analyses that will be conducted. Finally, the analysis plan should include the results of power and sample size calculations.

If there are separate analysis plans for the clinical and economic evaluations, efforts should be made to make them as consistent as possible to avoid real or perceived biases and increase transparency and consistency between studies (e.g., shared use of an intention-to-treat analysis, shared use of statistical tests for variables used commonly by both analyses, etc.). At the same time, the outcomes of the clinical and economic studies can differ (e.g., the primary outcome of the clinical evaluation might focus on event-free survival while the primary outcome of the economic evaluation might focus on quality-adjusted survival). Thus, the two plans need not be identical.

The analysis plan should also indicate the level of blinding that will be imposed on the analyst. Most, if not all, analytic decisions should be made while the analyst is blinded to the treatment groups. It is preferable that analysts have no variable (data) that represents treatment group assignment, including variables with blinded labels such as "treatment A" or "treatment B," even if the analyst has no knowledge about the actual treatments represented by the labels. Blinding is particularly important when investigators have not precisely specified the models that will be estimated, but instead rely on the structure of the data to help make decisions about these issues.

Methods for analysis of costs

When one analyzes cost data derived from randomized trials, one should report means of costs for the groups under study as well as the difference in the means, measures of variability and precision, such as the standard deviation and quantiles of costs (particularly if the data are skewed), and an indication of whether or not the costs are likely to be meaningfully different from each other in economic terms.

Traditionally, the determination of a difference in costs between the groups has been made using Student's t-tests or analysis of variance (ANOVA) (univariate analysis) and ordinary least squares regression (multivariable analysis). More recently analysts have moved toward the use of generalized linear models to improve the predictive power of multivariable analyses.[102]

Univariate analysis

A basic assumption underlying t-tests and ANOVA (which are parametric tests) is that cost data are normally distributed. Given that the distribution of these data often violates this assumption, a number of analysts have begun using non-parametric tests, such as the Wilcoxon rank-sum test (a test of median costs) and the Kolmogorov–Smirnov test (a test for differences in cost distributions), which make no assumptions about the underlying distribution of costs. The principal problem with these non-parametric approaches is that statistical conclusions about the mean need not translate into statistical conclusions about the median (e.g., the means could differ yet the medians could be identical), nor do conclusions about the median necessarily translate into conclusions about the mean. Similar difficulties arise when—to avoid the problems of non-normal distribution—one analyzes cost data that have been transformed to be more normal in their distribution (e.g., the log transformation of the square root of costs).

Table 38.2 shows the results of the univariate analysis of hospital costs measured among men receiving vehicle and an investigational medication for the treatment of aneurysmal subarachnoid hemorrhage.[103] The mean cost for patients receiving vehicle was $20 287 (SD, $22 542); the mean cost for patients receiving the investigational medication was $25 185 (SD, $22 619). The distribution (as seen from the quantiles reported in Table 38.2, which shows the distribution of costs for the two groups) is skewed. For example, the difference between the 25th and 50th percentiles

Table 38.2 Hospital costs of tirilazad mesylate for subarachnoid hemorrhage in men

Variable	Vehicle	Tirilazad 6 mg/kg per day
Cost ($)	20 287	25 185
Standard deviation	(22 542)	(22 619)
Distribution		
5%	4 506	10 490
25%	9 691	13 765
50%	13 773	18 834
75%	23 044	31 069
95%	53 728	51 771
Comparison of differences		
t-test	0.15	
t-test (log of costs)	0.02	
Wilcoxon rank-sum	0.001	
Kolmogorov–Smirnov	0.001	

is approximately $4500 for the two treatment groups, but is approximately $10 000 between the 50th and 75th percentiles. Of note, from the 5th to the 75th percentile, there was approximately a $5000 difference between the two treatment groups. By the 95th percentile, the costs in the two groups were similar. These distributions provide evidence that the costs differ between the two treatment groups.

The parametric and non-parametric statistical tests, however, yielded conflicting conclusions about whether or not the cost differences were statistically different from one another. The t-test comparing mean costs between the groups indicated a non-significant difference ($p = 0.15$), whereas the t-test comparing the mean log of costs and both of the non-parametric statistical tests indicated they differed ($p < 0.02$). In this case, one might conclude that the difference in the medians between groups is statistically significant, whereas the difference in the means between groups is not. Similarly conflicting conclusions about the statistical significance of observed differences in costs have been reported in other studies.[104] Although each of these statistical tests is informative, given that the important outcome for the analysis of the value for the cost of the new therapy (e.g., the cost-effectiveness ratio) is the difference in mean costs, the statistical test of differences in means (e.g., t-test) should be used for inferences about this outcome. Measuring the correct parameter should take precedence over threats to the efficiency of the way that parameter is measured.

Multivariable analysis

Regression analysis often is used to assess differences in costs, in part because the sample size needed to detect economic differences may be larger than the sample needed to detect clinical differences (i.e., to overcome power problems). Traditionally, ordinary least squares regression has been used to predict costs (or their log) as a function of the treatment group while controlling for covariables such as disease severity, costs prior to randomization, etc. However, use of the log of costs as the outcome variable simply to avoid statistical problems posed by untransformed costs leaves one with the problem that we are not interested in this outcome itself; rather we are interested in the difference in untransformed costs. In addition, the retransformation of the predicted difference in the log of costs into an estimate of the predicted difference in costs is not trivial.[105,106]

While univariate t-tests and ANOVAs assume the normal distribution of cost data, ordinary least squares regression assumes that the error terms from the prediction of costs are normally distributed. Because of the potential violation of this assumption, however, a number of alternative multivariable methods have been proposed for analyzing costs. In addition to the generalized linear model mentioned above, these methods include non-parametric hazards models,[107–111] parametric failure-time models,[107] Cox semiparametric regression,[112] and joint distributions of survival and cost.[113] The relative merits of several of these methods have been compared by Lipscomb and colleagues[114] and by Manning and Mullahy.[115] However, there is little conclusive evidence regarding which model is best in a given analytic circumstance.

More recently, generalized linear models have been adopted to address analytic issues associated

Table 38.3 Selected coefficients and *p* values for the hospital cost regressions for men receiving tirilazad for subarachnoid hemorrhage

	Coefficient	*p*
Intercept	1747	0.90
Randomization group*		0.05
6 mg/kg per day	6058	
2 mg/kg per day	−100	
0.6 mg/kg per day	−247	
Neurograde of subarachnoid hemorrhage		0.0001
Grade II	3950	
Grade III	3904	
Grade IV	9132	
Grade V	5406	

*6 mg/kg/day versus vehicle, 2 mg/kg/day, and 0.6 mg/kg/day, $p = 0.03$, 0.03, and 0.02, respectively; no other comparisons statistically significant.

with skewed and heteroscedastic cost distributions and to overcome issues related to retransformation.[116] Generalized linear models are prevalent in the literature and are typically specified with log links and gamma distributions to model the error term.[102,117,118] However, these specifications are not always appropriate, and an alternative approach, "extended estimating equations," can be applied whereby the functional form is derived empirically.[119]

Table 38.3 shows selected results of an ordinary least squares regression predicting hospital costs measured among men receiving vehicle and the investigational medication for the treatment of aneurysmal subarachnoid hemorrhage. On average, costs among those receiving the investigational medication were $6058 higher than costs among patients receiving vehicle ($p = 0.03$). Increasing levels in the neurograde of subarachnoid hemorrhage upon entry to the study (grades of subarachnoid hemorrhage range from I to V, with V being the most severe) were generally associated with increasing costs; the reduction in costs among those in grade V was due principally to the large number of patients in this category who died in the hospital. Other predictors of hospital costs included the additional days between onset of subarachnoid hemorrhage and randomization into the trial (+), age (+), and country (+/−) (data not shown).[73]

Uncertainty in economic assessment

There are a number of sources of uncertainty surrounding the results of economic assessments. One source relates to sampling error (stochastic uncertainty). The point estimates are the result of a single sample from a population. If we ran the experiment many times, we would expect the point estimates to vary. One approach to addressing this uncertainty is to construct confidence intervals both for the separate estimates of costs and effects as well as for the resulting cost-effectiveness ratio. A substantial literature has developed related to construction of confidence intervals for cost-effectiveness ratios.[120–123]

A common method for deriving 95% confidence intervals for cost-effectiveness ratios is the nonparametric bootstrap method.[124] In this method, one re-samples from the study sample and computes cost-effectiveness ratios in each of the multiple samples. To do so, one (i) draws a sample of size *n* with replacement from the empiric distribution and uses it to compute a cost-effectiveness ratio; (ii) repeats this sampling and calculation of the ratio (by convention, at least 1000 times for confidence intervals); (iii) orders the repeated estimates of the ratio from "lowest" to "highest" (ordering is complicated when we observe negative ratios that represent the therapy's dominance and its being dominated or that represent the therapy costing more and being more effective as well as it costing less and being less effective); and (iv) identifies a 95% confidence interval from this rank-ordered distribution. The percentile method is one of the simplest means of identifying a confidence interval, but it may not be as accurate as other methods. When using 1000 repeated estimates, the percentile method uses the 26th and 975th ranked cost-effectiveness ratios to define the confidence interval.[125]

In the multivariable regression analysis above, we estimated that therapy with the investigational medication added $6058 to the cost of hospitalization (95% CI: $693 to $11 423). The

results of a logistic regression predicting death indicated that the investigational medication yielded a difference in the predicted probability of death of 0.225.[103] The cost per death averted was $26 924 ($6058/0.225). The results of the bootstrap analysis indicated that the 95% CI for the cost-effectiveness ratio ranged from $4300 to $54 600.[106] Interpreting the results of the bootstrap in a Bayesian sense, evaluating stochastic uncertainty alone, there is a 96% chance that the ratio is below $50 000 per death averted.

In addition to addressing stochastic uncertainty, one may want to address uncertainty related to parameters measured without variation (e.g., unit cost estimates, discount rates, etc.), whether or not the results are generalizable to settings other than those studied in the trial, and, for chronic therapies, whether the cost-effectiveness ratio observed within the trial is likely to be representative of the ratio that would have been observed if the trial had been conducted for a longer period. These sources of uncertainty are often addressed using sensitivity analysis.

Currently available solutions

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 set of case studies that illustrate the practical application of these methods to the evaluation of pharmaceuticals. The following cases illustrate cost-effectiveness analyses of valsartan for treatment of chronic heart failure, tirilazad mesylate for aneurysmal subarachnoid hemorrhage, and high-dose chemotherapy plus autologous stem cell transplantation for patients with metastatic breast cancer.

Multinational economic evaluation of valsartan in patients with chronic heart failure

In this study, data on resource use and direct medical costs were analyzed to assess the economic impact of an angiotensin receptor blocker, valsartan, in combination with prescribed standard therapy for patients with New York Heart Association class II to IV heart failure.[36,126] Patients who enrolled in this clinical trial were receiving a standard regimen of medication for heart failure (e.g., angiotensin-converting enzyme (ACE) inhibitors) and were randomized to receive 160 mg of valsartan or placebo twice daily. A total of 5010 patients in 16 countries were enrolled in the trial. The clinical investigators found no differences in mortality; however, the valsartan group had a lower risk of experiencing the combined mortality–morbidity end point (i.e., death, hospitalization for heart failure, cardiac arrest with resuscitation, or receipt of intravenous inotropic or vasodilator drugs). Most of the difference was attributable to a lower risk of first hospitalization for heart failure among patients receiving valsartan. Economic data were collected prospectively as part of the clinical trial.

Resource use data were collected in a case report form at regular trial visits every 2 weeks for 2 months, at 4 months and 6 months, and then every 3 months throughout the duration of the follow-up period. Unit costs were collected for each of the resource categories assessed (hospitalizations, outpatient visits, and medications). For US patients, the cost estimates for hospital resources and outpatient visits were based on 1999 Medicare reimbursement rates. Unit costs for medications for all countries were derived from an international drug pricing database. For patients in countries outside of the United States, local health economists in each country provided mean cost estimates of outpatient and hospital care for discharge diagnoses, including heart failure, acute myocardial infarction, and several others. In most cases, these unit cost estimates were based on national fee schedules or hospital accounting systems. Cost estimates were converted to US dollars using purchasing power parties from the Organization for Economic Cooperation and Development. Costs were reported in 1999 US dollars.

To assign costs to hospitalizations for which the diagnosis was not included in the cost survey, unit costs for individual countries were imputed using diagnosis related group (DRG) weights from the US Centers for Medicare and Medicaid Services and the cost estimates provided for hospitalizations due

to heart failure. At that point, country-level costs were assigned to individual hospital events and adjusted for differences in length of stay. Daily medication costs were assigned to patients based on a mean daily dose indicated for patients with heart failure and on the duration for which the patient was on the medication.

The results of the resource use analysis indicate that half of the patients in each group were hospitalized at least once during the trial; however, patients in the valsartan group were 13.9% less likely than patients in the placebo group to have a heart failure hospitalization. In addition, patients in the valsartan group spent less time in hospital than did patients in the placebo group.

The results of the cost analysis indicate that the mean cost of a hospitalization for heart failure was $423 less for patients in the valsartan group, compared to patients in the placebo group. However, much of the savings was offset by higher costs for non-heart-failure hospitalizations among patients in the valsartan group, yielding a non-significant decrease in inpatient costs of $193 for patients in the valsartan group. The difference in outpatient costs also was non-significant. Overall within-trial costs, including the cost of valsartan, were $545 higher for patients in the valsartan group.

In exploratory subgroup analyses, the investigators found that costs were higher for patients aged 65 years and older, and the difference in costs between treatment groups was greater among older patients. Also, costs varied according to the heart failure medications patients were taking at baseline. Patients receiving valsartan and who were not taking an ACE inhibitor at baseline had lower morbidity and mortality and $929 lower costs compared to their counterparts receiving placebo, even after including the cost of valsartan. Thus, valsartan was the dominant strategy in this subgroup of patients. However, patients receiving valsartan and who used both an ACE inhibitor and a beta-blocker at baseline had lower survival and higher costs relative to placebo. Thus, valsartan was the dominated strategy in this subgroup. Analysis of the subgroup of patients receiving valsartan and who used an ACE inhibitor without a beta-blocker at baseline was inconclusive.

The authors conclude that patients receiving valsartan experienced clinical benefit at a mean incremental cost of $285 per year. In patients who were not taking an ACE inhibitor at baseline, valsartan was the dominant strategy.

Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage

The investigators undertook a cost-effectiveness analysis on the use of tirilazad mesylate, a potential free-radical scavenger and lipid peroxidation inhibitor, for the treatment of aneurysmal subarachnoid hemorrhage in a randomized, double-blind, placebo-controlled Phase III clinical trial.[103] A sample of 1023 patients from nine European countries, Australia, and New Zealand were randomized to receive one of four treatments: vehicle or tirilazad 0.6, 2.0, or 6 mg/kg of body weight per day for 8 to 10 days. All treatments were administered within 48 hours of the occurrence of a subarachnoid hemorrhage and ceased 10 days after the initial hemorrhagic event. All patients received nimodipine treatment concomitantly.

Clinical and economic outcomes at 3 months and hospital costs were estimated using data from 1019 of the 1023 patients enrolled in the study. Death during the 3 months after randomization was the clinical outcome. The primary clinical outcome was occurrence of vasospasm, and the secondary outcome was Glasgow Outcome Scale (GOS) score and mortality. The authors evaluated the costs of hospitalizations during the trial as well as patients' residence and employment status 3 months after randomization. Cost estimates were based on resource use and unit costs for resources used.

Data on the length of hospital stay, number of imaging studies, number and types of surgical procedures, and medication use were collected prospectively. Information on the site of care in the hospital was obtained retrospectively but was collected before the study results were known and while the investigators were blinded to the treatment groups. Patient status at 3 months was evaluated prospectively by assessing daily residence costs for patients living at home with supervision or

dependent on others as well as for those in minimal care, skilled care, or long-term rehabilitation institutions. The daily employment value at 3 months was also assessed for homemakers and for full- and part-time workers.

Local health economists from six countries collected unit costs of inpatient resource utilization. The averages of the unit costs from the six countries were used for the five other countries. The $137.50 cost per 150 mg of tirilazad was based on a price set by the manufacturer. The authors determined values for employment using wage and salary data from the participating countries. The unit costs from other countries were converted into 1993 US dollars. The authors stated that during the sensitivity analysis, deaths averted were translated into gains in life expectancy both with and without adjustments for quality-of-life.

The results of the study indicate that patients were similar in all groups except in the proportions having right-to-left and left-to-right shifts of the midline structures and those having generalized, as opposed to localized, brain swelling. Total length of hospital stay, number of days between the onset of subarachnoid hemorrhage and randomization, number of days the patient was intubated, characteristics of the hemorrhage, the country in which patients received care, and mortality of the patients were all predictors of hospital stay by unit type.

Results of the economic analysis showed that the average hospital cost was $20 341 (SD, ±$17 239) for the whole sample. The average hospital cost for women ($19 569 ± $15 156) was less than the cost among men ($21 835 ± $20 743). The results also indicated that the majority of the cost was attributable to the length of stay and the greatest difference in cost was due to the costs of tirilazad. The cost analysis at 3 months showed that the largest difference in employment value was observed between men who received tirilazad 6 mg/kg per day and those who received vehicle ($9.20 additional earnings per day). In addition, the results showed that the largest difference in residence cost was also between these two groups ($15.80 additional residence cost per day for the 6 mg/kg group). However, none of these differences was statistically significant. One significant finding of this study was that those who received tirilazad 6 mg/kg per day had a significant reduction in the probability of death in the whole sample ($p = 0.002$) and in men ($p = 0.0001$). There was no significant difference in the probability of death among women between the group that received tirilazad 6 mg/kg per day and those that received vehicle. When costs and outcomes were compared, the results showed that in both the entire sample, and in men, tirilazad 6 mg/kg per day was associated with improved survival compared to vehicle, but also with increased hospital costs. The cost per death averted was $29 615 for the sample as a whole and $26 924 for men. There were no significant differences in costs or probability of survival of women in either the tirilazad 6 mg/kg per day or vehicle group.

The results were subjected to a sensitivity analysis, showing that the cost-effectiveness ratio (95% CI) between those in the entire sample who received tirilazad 6 mg/kg per day and those who received vehicle was $9189 per death averted due to tirilazad, adding hospital costs and mortality. The cost-effectiveness ratios among men (95% CI) ranged from $4300 to $54 600 per death averted. The sensitivity analysis also showed that in 68.8% of women, 6 mg/kg per day of tirilazad resulted in an increase in hospital costs and survival. Five percent experienced decreased costs and survival, 11.6% had decreased costs and increased survival, and 14.3% had increased costs and decreased survival. Another finding was that in the entire sample, the ratios of cost per year of life saved and cost per quality-adjusted year of life saved fell below $50 000 if survivors live on average 0.6 and 0.8 years respectively. For men, these ratios fell below $50 000 if survivors at the end of the trial live an average of 1.1 and 2.4 years. Among men, the ratio of cost per year of life saved did not fall below $27 500. Also, the ratio of the cost per quality-adjusted year of life saved did not fall under $36 400.

The economic analysis of this study showed that treatment with tirilazad mesylate is associated with a significant increase in survival and increase in the cost of care. The results also showed that the ratios of cost per death averted, cost per year of life saved, and cost per quality-adjusted year of life saved are favorable when compared to other interventions.

Economic evaluation of high-dose chemotherapy plus autologous stem cell transplantation for metastatic breast cancer

The Philadelphia Bone Marrow Transplant Group conducted a clinical trial to compare survival associated with high-dose chemotherapy plus autologous hematopoietic stem cell transplantation versus conventional-dose chemotherapy in women with metastatic breast cancer. Data on resource use and costs were collected as secondary end points of the study. Because the clinical trial found no significant differences in survival between the two treatment groups, the economic evaluation[127] would provide important additional information about the two therapies.

Because resource use data were not captured explicitly in the clinical case report form, the investigators abstracted the clinical trial records and oncology department flow sheets retrospectively to document the resources used by each patient. The abstraction process captured information about hospitalizations, medical procedures, medications, tests, and inpatient and outpatient physician services for each patient from the time of randomization through death or end of follow-up. Each patient's course of treatment and resource use was analyzed in four phases—randomization, treatment, progression, and remission. (The treatment phase for patients in the transplantation group was further divided into an inpatient phase and postdischarge phase.) Based on these clinical phases, the investigators grouped the patients into one of three clinical "trajectories." Patients in trajectory 1 went through all four clinical phases before the end of the study. Patients in trajectory 2 went through randomization, treatment, and immediately to progression. Patients in trajectory 3 went through randomization, treatment, and immediately to remission until the end of the study.

Daily costs for inpatient care in both treatment groups were assigned according to each patient's length of hospital stay. They were estimated using data from the cost accounting system of an academic medical center. Cost estimates for transplantation hospitalizations were based on the mean daily cost of hospitalization for stem cell transplantation. Cost estimates for other hospitalizations were based on the mean daily cost of a hospitalization for neutropenic fever. The investigators used Medicare reimbursement rates to estimate costs for inpatient and outpatient laboratory tests and physician fees. They estimated medication costs by referring to each drug's average wholesale price. The study medication cost estimates were added to the estimate of inpatient costs for patients undergoing transplantation, and they were added to the estimate of outpatient costs for patients in the conventional-dose chemotherapy group. When a patient was missing costs for any month in the study, the investigators imputed the costs using median costs for each clinical phase, clinical trajectory, and treatment group.

The results of the economic analysis showed that patients in the transplantation group had significantly more inpatient days (28.6 versus 17.8; $p = 0.004$) and significantly greater mean length of stay per hospitalization (21.9 days versus 15.2 days; $p = 0.02$) than did patients in the conventional-dose chemotherapy group. Patients in the transplantation group also had more procedures per outpatient visit. Mean total costs were higher for patients in the transplantation group ($84 055 versus $28 169), for a mean cost difference of $55 886 (95% CI: $47 298 to $63 666). Most of the difference was attributable to the $52 448 difference in inpatient care. The investigators also found differences by clinical trajectory, and these differences were not consistent between treatment strategies. For example, outpatient costs for patients in trajectory 3 who were randomized to conventional-dose chemotherapy were much higher than outpatient costs for patients in trajectories 1 and 2. Because patients in trajectory 3 completed more cycles of treatment, they spent more time in the treatment phase and accrued greater costs associated with administering therapy.

In sensitivity analysis, the investigators confirmed the robustness of their findings by varying the discount rate, the hospital costs, and the number of cycles of paclitaxel and docetaxel that patients were assumed to have received. Varying the discount rate had little effect on the mean difference in cost between treatment groups.

Increasing and decreasing the hospital costs by 50% yielded mean differences in total costs ranging from $36 528 to $75 531. Increasing the number of cycles of paclitaxel and docetaxel caused a greater increase in costs for patients in the conventional-dose chemotherapy group, because more patients in this group were treated with these drugs.

The authors concluded that high-dose chemotherapy plus stem-cell transplantation for women with metastatic breast cancer was more costly and resulted in greater morbidity with no improvement in survival. By studying resource use and estimating costs, the authors were able to quantify the economic burden associated with the two treatments and to provide information about the clinical trajectories of patients with metastatic breast cancer.

The future

The emergence of cost as a criterion for the evaluation of pharmaceutical products requires the continued development and application of research methods to guide decision makers. Patients, and physicians acting on their behalf, are principally concerned about the effectiveness and safety of drugs. However, as patients, payers, and society become more concerned about the cost of medical care, the clinical contribution of pharmaceutical agents will be weighed against their costs and compared with the next best alternative. As third-party payers increasingly cover drug costs, they will be concerned with their expenditures on pharmaceuticals and the value obtained for the money spent. Hospitals and other providers of care, operating under increasingly constrained budgets, will increase their assessments of pharmaceutical expenditures. In the United States, comparative effectiveness research and evidenced-based medicine are likely to shape access to clinical therapies (see Chapter 32).

The naive decision maker might weigh drugs according to their purchase price alone. This paradigm ignores two essential elements in choosing pharmaceuticals. First, in identifying a drug's cost, its purchase price is only part of its real economic impact. The costs of preparation and delivery, as well as the cost of monitoring for and treating adverse events and side effects, are unavoidable elements of the cost of treating patients.

Second, a full analysis should go beyond the identification of cost. Only if the safety and effectiveness of two pharmaceutical agents are equivalent will cost alone determine the choice of therapy. Cost-effectiveness analysis requires that cost be weighed against effectiveness and that when two or more alternatives are being compared, the additional cost per additional unit of effectiveness be measured. Beyond these considerations of cost identification and cost-effectiveness, a full economic analysis will also assess the net value, or utility, of the drug's clinical contribution. Moreover, cost-effectiveness is about relative value in specific therapies. Decision makers under austere budgets may need to consider whether therapies are cost saving if they want to maintain a fixed health-care budget.

This is a challenging period for the field of clinical economics. Many of the earlier methodologic challenges of the field have been addressed, and researchers have gained experience in implementing economic evaluations in a multitude of settings. This experience has raised new questions for those interested in the development of new clinical therapies and in the application of economic data to the decision making process.

With the increasing importance of multinational clinical trials in the clinical development process, many of the problems facing researchers today involve the conduct of economic evaluations in multinational settings. Foremost among these is the problem of generalizability.[38] There is little consensus among experts as to whether the findings of multinational clinical trials are more generalizable than findings from trials conducted in single countries. This question is even more problematic for multinational economic evaluations, because the findings of economic evaluations reflect complex interactions between biology, epidemiology, practice patterns, and costs that differ from country to country.[128]

As physicians are asked simultaneously to represent their patients' interests while being asked to

deliver 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,[129] 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 authors dedicate this chapter to John M. Eisenberg, MD (1946–2002), mentor, friend, and fellow author in previous editions of this work.

The authors also acknowledge Harris Koffer, Pharm D, vice president of Quest Diagnostics, Inc., for his contributions to this chapter.

References

1. Bombardier C, Eisenberg J. Looking into the crystal ball: can we estimate the lifetime cost of rheumatoid arthritis? *J Rheumatol* 1985; **12**: 201–4.
2. Drummond MF. *Principles of Economic Appraisal in Health Care*. New York: Oxford University Press, 1980.
3. 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.
4. Drummond MF, Davies L. Economic analysis alongside clinical trials: revisiting the methodological issues. *Int J Technol Assess Health Care* 1991; **7**: 561–73.
5. Detsky AS, Naglie IG. A clinician's guide to cost effectiveness analysis. *Ann Intern Med* 1990; **113**: 147–54.
6. Eddy DM. Cost-effectiveness analysis: is it up to the task? *JAMA* 1992; **267**: 3342–8.
7. 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**: S909.
8. Eisenberg JM, Koffer H, Glick HA, Connell ML, Loss LE, Talbot GH, *et al*. What is the cost of nephrotoxicity associated with aminoglycosides? *Ann Intern Med* 1987; **107**: 900–9.
9. Eisenberg JM. Clinical economics: a guide to the economic analysis of clinical practices. *JAMA* 1989; **262**: 2879–86.
10. 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.
11. Warner KE, Luce BR. *Cost Benefit and Cost Effectiveness Analysis in Health Care: Principles, Practice and Potential*. Ann Arbor, MI: Health Administration Press, 1982.
12. Weinstein MC. Economic assessments of medical practices and technologies. *Med Decis Making* 1981; **1**: 309.
13. Weinstein MC, Stason WB. Cost-effectiveness of coronary artery bypass surgery. *Circulation* 1982; **66**: 56–68.
14. Weinstein MC. Principles of cost-effective resource allocation in health care organizations. *Int J Technol Assess Health Care* 1990; **6**: 93–103.
15. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
16. Glennie JL, Torrance GW, Baladi JF, Berka C, Hubbard E, Menon D, *et al*. The revised Canadian guidelines for the economic evaluation of pharmaceuticals. *Pharmacoeconomics* 1999; **15**: 459–68.
17. Commonwealth Department of Human Services and Health. *Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee Including Major Submission Involving Economic Analyses*. Canberra, Australia: Australian Government Publishing Service, 1995.
18. Drummond MF. The use of health economic information by reimbursement authorities. *Rheumatology* 2003; **42** (Suppl 3): iii60–iii3.
19. Nishimura S, Torrance GW, Ikegami N, Fukuhara S, Drummond M, Schubert F. Information barriers to the implementation of economic evaluations in Japan. *Pharmacoeconomics* 2002; **20** (Suppl 2): 9–15.
20. National Institute for Clinical Guidance. *Guidance for Manufacturers and Sponsors*, Technology Appraisals Process Series no. 5. London: National Institute for Clinical Excellence, 2001.
21. Pharmaceutical Management Agency Ltd (PHARMAC). *A Prescription for Pharmacoeconomic Analysis*, version 1, September 24, 1999.

22. Elsinga E, Rutten FF. Economic evaluation in support of national health policy: the case of the Netherlands. *Soc Sci Med* 1997; **45**: 605–20.
23. Szende Á, Mogyorósy Z, Muszbek N, Nagy J, Pallos G, Dózsa D. Methodological guidelines for conducting economic valuation of healthcare interventions in Hungary: a Hungarian proposal for methodology standards. *Eur J Health Econ* 2002; **3**: 196–206.
24. Graf von der Schulenburg JM, Greiner W, Jost F, Klusen N, Kubin M, Leidl R, *et al*. German recommendations on health economic evaluation: third and updated version of the Hanover Consensus. *Value Health* 2008; **11**: 540–1.
25. National Institute for Health and Clinical Excellence. *Guide to the Methods of Technology Appraisal*, June 2008. http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (accessed July 2011).
26. Wibulpolprasert S; Subcommittee for Development of the National List of Essential Drugs. The need for guidelines and the use of economic evidence in decision-making in Thailand: lessons learnt from the development of the national list of essential drugs. *J Med Assoc Thai* 2008; **91** (Suppl 2): S1–3.
27. van Veledn ME, Severens JL, Novak A. Economic evaluations of healthcare programmes and decision making: the influence of economic evaluations on different healthcare decision-making levels. *Pharmacoeconomics* 2005; **23**: 1075–82.
28. Cholmenero F, Sullivan SD, Palmer JA, Brauer CA, Bungay K, Watkins J, Neumann P. Quality of clinical and economic evidence in dossier formulary submissions. *Am J Manag Care* 2007; **13**: 401–7.
29. Schulman KA, Kinosian BP, Jacobson TA, Glick HA, Willian MK, Koffer H, *et al*. Reducing high blood cholesterol level with drugs: cost effectiveness of pharmacologic management. *JAMA* 1990; **264**: 3025–33.
30. Schulman KA, Glick HA, Rubin H, Eisenberg JM. Cost effectiveness of HA-1A monoclonal antibody for gram-negative sepsis: prospective economic assessment of a new therapeutic agent. *JAMA* 1991; **226**: 3466–71.
31. 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**: 435–46.
32. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, *et al*. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health* 2005; **8**: 521–33.
33. Mark DB, Hlatky MA, Califf RM, Naylor CD, Lee KL, Armstrong PW, *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.
34. 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**: 1537–42.
35. 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**: 2574–83.
36. 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**: 122–8.
37. Reed SD, Radeva JI, Weinfurt KP, McMurray JJV, Pfeffer MA, Velazquez EJ, *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**: 323–9.
38. Reed SD, Anstrom KJ, Bakhai A, Briggs AH, Califf RM, Cohen DJ, *et al*. Conducting economic evaluations alongside multinational clinical trials: toward a research consensus. *Am Heart J* 2005; **149**: 434–43.
39. Hlatky MA, Owens DK, Sanders GD. Cost-effectiveness as an outcome in randomized clinical trials. *Clin Trials* 2006; **3**: 543–51.
40. 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**: 1624–32.
41. 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**: 1103–9.
42. 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**: 332–42.

43. Towse A, Pritchard C. What is NICE's 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.
44. Finkler SA. The distinction between cost and charges. *Ann Intern Med* 1982; **96**: 102–9.
45. Bridges JM, Jacobs P. Obtaining estimates of marginal cost by DRG. *Healthc Financ Manage* 1986: **40**: 40–6.
46. Hsiao WC, Braun P, Becker ER, Becker ER, Yntema D, Verrilli DK, *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): NS1–12.
47. Granneman TW, Brown RS, Pauly MV. Estimating hospital costs. *J Health Econ* 1986; **5**: 107–27.
48. 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**: 396–406.
49. Landefeld JS, Seskin EP. The economic value of life: linking theory to practice. *Am J Public Health* 1982; **72**: 555–65.
50. Rice DP, Hodgson TA. The value of human life revisited. *Am J Public Health* 1982; **72**: 536–8.
51. Thompson MS, Read JL, Liang M. Feasibility of willingness-to-pay measurement in chronic arthritis. *Med Decis Making* 1984; **4**: 195–215.
52. Mishan EJ. *Cost–Benefit Analysis*, 3rd edn. London: George Allen & Unwin, 1992.
53. Shelling TC, Chase SB, eds. The life you save may be your own. In: *Problems in Public Expenditure Analysis*. Washington, DC: The Brookings Institution, 1968.
54. 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**: 632–8.
55. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. *JAMA* 2003; **289**: 3135–44.
56. 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**: 2443–54.
57. Strosberg MA, Wiener JM, Baker R, Fein IA, eds. *Rationing America's Medical Care: The Oregon Plan and Beyond*. Washington, DC: The Brookings Institution, 1992.
58. Sulmasy DP. Physicians, cost-control, and ethics. *Ann Intern Med* 1992; **116**: 920–6.
59. Glick HA, Doshi JA, Sonnad SS, Polsky D. *Economic Evaluation in Clinical Trials*. Oxford: Oxford University Press, 2007.
60. Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Malik F, *et al*. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. *Value Health* 2009; **12**: 409–18.
61. Evans WE, McLeod HL. Pharmacogenomics—drug disposition, drug targets, and side effects. *N Engl J Med* 2003; **348**: 538–49.
62. Gold HT, Hall MJ, Blinder V, Schackman BR. Cost effectiveness of pharmacogenetic testing for uridine diphosphate glucuronosyltransferase 1A1 before irinotecan administration for metastatic colorectal cancer. *Cancer* 2009; **115**: 3858–67.
63. 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**: 1624–32.
64. Laska EM, Meisner M, Seigel C. Power and sample size in cost-effectiveness analysis. *Med Decis Making* 1999; **19**: 339–43.
65. Willan AR. Analysis, sample size, and power for estimating incremental net health benefit from clinical trial data. *Control Clin Trials* 2001; **22**: 228–37.
66. Brown M, Glick HA, Harrell F, Herndon J, McCabe M, Moinpour C, *et al*. Integrating economic analysis into cancer clinical trials: the National Cancer Institute–American Society of Clinical Oncology Economics workbook. *J Natl Cancer Inst Monogr* 1998; **24**: 1–28.
67. Cannon CP, Harrington RA, James S, Ardissino D, Becker RC, Emanuelsson H, *et al*. Comparison of ticagrelor with clopidogrel in patients with a planned invasive strategy for acute coronary syndromes (PLATO): a randomised double-blind study. *Lancet* 2010; **375**: 283–93.
68. Cook JR, Drummond M, Glick H, Heyse JF. Assessing the appropriateness of combining economic data from multinational clinical trials. *Stat Med* 2003; **22**: 1955–76.
69. Drummond M, Pang F. Transferability of economic evaluation results. In: Drummond M, McGuire A, eds. *Economic Evaluation in Health Care. Merging Theory with Practice*. Oxford: Oxford University Press, 2001, pp. 254–76.
70. Willke RJ, Glick HA, Polsky D, Schulman K. Estimating country-specific cost-effectiveness from

multinational clinical trials. *Health Econ* 1998; **7**: 481–93.
71. O'Brien BS. A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. *Am J Manag Care* 1997; **3**: S33–40.
72. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, *et al*. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess* 2004; **8**: iii–iv, 1–192.
73. Willan AR, Pinto EM, O'Brien BJ, *et al*. Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. *Health Econ* 2005; **14**: 327–38.
74. Pinto EM, Willan AR, O'Brien BJ. Cost-effectiveness analysis for multinational clinical trials. *Stat Med* 2005; **24**: 1965–82.
75. Manca A, Rice N, Sculpher MJ, Briggs AH. Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Econ* 2005; **14**: 471–85.
76. Grieve R, Nixon R, Thompson SG, Cairns J. Multilevel models for estimating incremental net benefits in multinational studies. *Health Econ* 2007; **16**: 815–26.
77. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ* 2002; **11**: 415–30.
78. Sculpher MJ, Claxton KP, Drummond MF, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Econ* 2006; **15**: 677–87.
79. Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. *JAMA* 1998; **279**: 1278–81.
80. Wennberg JE, Fisher ES, Stukel TA, Skinner JS, Sharp SM, Bronner KK. Use of hospitals, physician visits, and hospice care during last six months of life among cohorts loyal to highly respected hospitals in the United States. *BMJ* 2004; **328**: 607.
81. Baker LC, Fisher ES, Wennberg JE. Variations in hospital resource use for Medicare and privately insured populations in California. *Health Aff (Millwood)* 2008; **27**: w123–34.
82. Anderson KM, Odell PM, Wilson PW, Kunnel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991; **121**: 293–8.
83. Stokes J 3rd, Kannel WB, Wolf PA, Dagostino RB, Lupples LA. Blood pressure as a risk factor for cardiovascular disease. The Framingham Study—30 years of follow-up. *Hypertension* 1989; **13** (Suppl): I13–18.
84. Schulman KA, Lynn LA, Glick HA, Eisenberg JM. Cost effectiveness of low-dose zidovudine therapy for asymptomatic patients with human immunodeficiency virus (HIV) infection. *Ann Intern Med* 1991; **114**: 798–802.
85. Glick H, Heyse JF, Thompson D, Epstein RS, Smith ME, Oster G. A model for evaluating the cost-effectiveness of cholesterol-lowering treatment. *Int J Technol Assess Health Care* 1992; **8**: 719–34.
86. 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**: 1145–51.
87. Grover SA, Abrahamowicz M, Joseph L, Brewer C, Coupal L, Suissa S. The benefits of treating hyperlipidemia to prevent coronary heart disease: expected changes in life expectancy and morbidity. *JAMA* 1992; **267**: 816–22.
88. Mark DB, Talley JD, Topol EJ, Bowman L, Lam LC, Anderson KM, *et al*. Economic assessment of platelet glycoprotein IIb/IIIa inhibition for prevention of ischemic complications of high-risk coronary angioplasty. EPIC Investigators. *Circulation* 1996; **94**: 629–35.
89. Lincoff AM, Mark DB, Tcheng JE, Califf RM, Bala MV, Anderson KM, *et al*. Economic assessment of platelet glycoprotein IIb/IIIa receptor blockade with abciximab and low-dose heparin during percutaneous coronary revascularization: results from the EPILOG randomized trial. Evaluation in PTCA to improve long-term outcome with abciximab GP IIb/IIIa blockade. *Circulation* 2000; **102**: 2923–9.
90. Topol EJ, Mark DB, Lincoff AM, Cohen E, Burton J, Kleiman N, *et al*. Outcomes at 1 year and economic implications of platelet glycoprotein IIb/IIIa blockade in patients undergoing coronary stenting: results from a multicentre randomised trial. EPISTENT Investigators. Evaluation of platelet IIb/IIIa inhibitor for stenting. *Lancet* 1999; **354**: 2019–24.
91. Drummond MF, Becker DL, Hux M, Chancellor JV, Duprat-Lomon I, Kubin R, *et al*. An economic evaluation of sequential i.v./ po moxifloxacin therapy compared to i.v./ po co-amoxiclav with or without clarithromycin in the treatment of community-acquired pneumonia. *Chest* 2003; **124**: 526–35.
92. Lorber MI, Fastenau J, Wilson D, DiCesare J, Hall ML. A prospective economic evaluation of basilixi-

mab (Simulect) therapy following renal transplantation. *Clin Transplant* 2000; **14**: 479–85.

93. Ramsey SD, Moinpour CM, Lovato LC, Crowley JJ, Grevstad P, Presant CA, *et al*. Economic analysis of vinorelbine plus cisplatin versus paclitaxel plus carboplatin for advanced non-small-cell lung cancer. *J Natl Cancer Inst* 2002; **94**: 291–7.
94. Reed SD, Li Y, Anstrom KJ, Schulman KA. Cost effectiveness of ixabepilone plus capecitabine for metastatic breast cancer progressing after anthracycline and taxane treatment. *J Clin Oncol* 2009; **27**: 2185–91.
95. Neumann PJ, Fang CH, Cohen JT. 30 years of pharmaceutical cost-utility analyses: growth, diversity and methodological improvement. *Pharmacoeconomics* 2009; **27**: 861–72.
96. Polsky D, Doshi JA, Bauer MS, Glick HA. Clinical trial-based cost-effectiveness analyses of antipsychotic use. *Am J Psychiatry* 2006; **163**: 2047–56.
97. Greenberg D, Earle C, Fang CH, Eldar-Lissai A, Neumann PJ. When is cancer care cost-effective? A systematic overview of cost-utility analyses in oncology. *J Natl Cancer Inst* 2010; **102**: 82–8.
98. Ladapo JA, Neumann PJ, Keren R, Prosser LA. Valuing children's health: a comparison of cost-utility analyses for adult and paediatric health interventions in the US. *Pharmacoeconomics* 2007; **25**: 817–28.
99. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Med Care* 1992; **30**: 231–43.
100. Hillman AL, Eisenberg JM, Pauly MV, Bloom BS, Glick H, Kinosian B, *et al*. Avoiding bias in the conduct and reporting of cost-effectiveness research sponsored by pharmaceutical companies. *N Engl J Med* 1991; **324**: 1362–5.
101. Udvarhelyi IS, Colditz GA, Ri A, Epstein AM. Cost-effectiveness and cost–benefit analyses in the medical literature: are the methods being used correctly? *Ann Intern Med* 1992; **116**: 238–44.
102. Polsky D, Glick H. Costing and cost analysis in randomized controlled trials: caveat emptor. *Pharmacoeconomics* 2009; **27**: 179–88.
103. Glick H, Willke R, Polsky D, Llana T, Alves WM, Kassell N, *et al*. Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage: economic evaluation of a phase III clinical trial in Europe and Australia. *Int J Technol Assess Health Care* 1998; **14**: 145–60.
104. Zhou X-H, Melfi CA, Hui SL. Methods for comparison of cost data. *Ann Intern Med* 1997; **127**: 752–6.
105. Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. *J Health Econ* 1998; **17**: 283–95.
106. Duan N. Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc* 1983; **78**: 605–10.
107. Dudley RA, Harrell FE, Smith LR, Mark DB, Califf RM, Pryor DB, *et al*. Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. *J Clin Epidemiol* 1993; **46**: 261–71.
108. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997; **53**: 113–28.
109. Fenn P, McGuire A, Phillips V, Backhouse M, Jones D. The analysis of censored treatment cost data in economic evaluation. *Med Care* 1995; **33**: 851–61.
110. Fenn P, Mcguire A, Backhouse M, Jones D. Modeling programme costs in economic evaluation. *J Health Economics* 1996; **15**: 115–25.
111. Schulman KA, Buxton M, Glick H, Sculpher M, Guzman G, Kong J, *et al*. Results of the economic evaluation of the FIRST study. *Int J Technol Assess Health Care* 1996; **12**: 698–713.
112. Harrell F, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; **15**: 361–87.
113. Lancaster T, Intrator O. *Panel Data with Survival: Hospitalization of HIV Patients*. Brown University Department of Economics, Working Paper Series, #95–36, 1995.
114. Lipscomb J, Ancukiewicz M, Parmigiani G, Hasselblad V, Samsa G, Matchar DBJ. Predicting the cost of illness: a comparison of alternative models applied to stroke. *Med Decis Making* 1998; **18**: S39–56.
115. Manning WG, Mullahy J. Estimating log models: to transform or not to transform. *J Health Econ* 2000; **20**: 461–94.
116. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. *J Health Econ* 2005; **24**: 465–88.
117. Blough DK, Madden CW, Hornbrook MC. Modeling risk using generalized linear models. *J Health Econ* 1999; **18**: 153–71.
118. Moran JL, Solomon PJ, Peisach AR, Martin J. New models for old questions: generalized linear models for cost prediction. *J Eval Clin Pract* 2007; **13**: 381–9.
119. Basu A, Arondekar BV, Rathouz PJ. Scale of interest versus scale of estimation: comparing alternative

estimators for the incremental costs of a comorbidity. *Health Econ* 2006; **15**: 1091–107.

120. Heyse J, Cook J. *A New Measure of Cost Effectiveness in Comparative Clinical Economic Trials*. Boston, MA: American Statistical Association, 1992.
121. O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost effectiveness studies in health care. *Med Care* 1994; **32**: 150–63.
122. Willan AR, O'Brien BJ. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Econ* 1996; **5**: 297–305.
123. Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Stat Med* 1996; **15**: 1447–58.
124. Polsky DP, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Econ* 1997; **6**: 243–52.
125. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
126. Cohn JN, Tognoni G, Valsartan Heart Failure Trial Investigators. A randomized trial of the angiotensin-receptor blocker valsartan in chronic heart failure. *N Engl J Med* 2001; **345**: 1667–75.
127. Schulman KA, Stadtmauer EA, Reed SD, Glick HA, Goldstein LJ, Pines JM, *et al*. Economic analysis of conventional-dose chemotherapy compared with high-dose chemotherapy plus autologous hematopoietic stem-cell transplantation for metastatic breast cancer. *Bone Marrow Transplant* 2003; **31**: 205–10.
128. Glickman SW, McHutchison JG, Peterson ED, Cairns CB, Harrington RA, Califf RM, *et al*. Ethical and scientific implications of the globalization of clinical research. *N Engl J Med* 2009; **360**: 816–23.
129. Eisenberg JM. From clinical epidemiology to clinical economics. *J Gen Intern Med* 1988; **3**: 299–300.

CHAPTER 39

Using Quality-of-Life Measurements in Pharmacoepidemiologic Research

Holger Schünemann,[1,2] *Bradley C. Johnston,*[1] *Roman Jaeschke,*[1,2] *and Gordon H. Guyatt*[1,2]

[1] Department of Clinical Epidemiology and Biostatistics, McMaster University, Health Sciences Center, Hamilton, Ontario, Canada
[2] Department of Medicine, St Joseph's Hospital, Hamilton, Ontario, Canada

Introduction

One may judge the impact of drug interventions by examining a variety of outcomes. In some situations, the most compelling evidence of drug efficacy may be found as a reduction in mortality (beta-blockers after myocardial infarction), rate of hospitalization (neuroleptic agents for schizophrenia), rate of disease occurrence (antihypertensives for strokes), or rate of disease recurrence (chemotherapy after surgical cancer treatment). Alternatively, clinicians frequently rely on direct physiological or biochemical measures of the severity of a disease process and the way drugs influence these measures—for example, left ventricular ejection fraction in congestive heart failure, spirometry in chronic airflow limitation, or glycosylated hemoglobin level in diabetes mellitus.

However, clinical investigators have recognized that there are other important aspects of the usefulness of the interventions which these epidemiologic, physiologic, or biochemical outcomes do not address, and are typically patient-reported outcomes. These areas encompass the ability to function normally; to be free of pain and physical, psychological, and social limitations or dysfunction; and to be free from iatrogenic problems associated with treatment. On occasion, the conclusions reached when evaluating different outcomes may differ: physiologic measurements may change without people feeling better,[1,2] a drug may ameliorate symptoms without a measurable change in physiologic function, or life prolongation may be achieved at the expense of unacceptable pain and suffering.[3] The recognition of these patient-important (versus disease-oriented) and patient-reported areas of well-being led to the introduction of a technical term: health-related quality-of-life (HRQL).

The term "quality-of-life," as it is often used, lacks focus and precision and, because it is an abstract concept, its definition has led to much debate. Since the patient's subjective well-being is influenced by many factors unrelated to the disease process or treatment (e.g., education, income, quality of the environment, etc.), investigators have adopted the narrower term, HRQL. Some definitions of HRQL reflect the evaluation of patients' overall well-being in several broad domains (physiologic, functional, psychological, and social status), and in subcomponents of each domain (e.g., pain, sleep, activities of daily living, and sexual function within physical and functional domains).

It follows that HRQL is a multifactorial concept that, from the patient's perspective, represents the final common pathway of all the physiological, psychological, and social influences of the therapeutic

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

process.[4] It follows also that, when assessing the impact of a drug on a patient's HRQL, one may be interested in describing the patient's status (or changes in the patient status) on a whole variety of domains, and that different strategies and instruments are required to explore separate domains.

Definitions of HRQL, both theoretical and practical, remain controversial. Most HRQL measurement instruments focus largely on how patients are functioning, e.g., their ability to care for themselves and carry out their usual roles in life. While this pragmatic view of HRQL has gained ascendancy, there remain those who argue that, unless you are tapping into individual patients' own values and preferences of health states, you may be measuring health status but you are not measuring HRQL.[5] Consider, for instance, a woman with quadriplegia who, despite her limitations, is very happy and fulfilled and values her life highly (more, for instance, than most people, or more than she did before she suffered quadriplegia). On most domains of most HRQL instruments, this woman's results would suggest a poor HRQL, despite the high value she places on her health state. Investigators and those interpreting the results of HRQL measure should be aware of the varying emphasis put on individual patient values and preferences in the different types of instruments.[6]

Clinical problems to be addressed by pharmacoepidemiologic research

HRQL effects may be pertinent in investigating and documenting both beneficial as well as harmful aspects of drug action. The knowledge of these drug effects may be important, not only to the regulatory agencies and physicians prescribing the drugs, but to the people who agree to take the medication and live with both its beneficial actions and detrimental side effects. Investigators must therefore recognize the clinical situations where a drug may have an important effect on HRQL. This requires careful examination of data available from earlier phases of drug testing and, until now, has usually been performed in the latter stages of Phase III testing. For example, Croog and colleagues studied the effect of three established antihypertensive drugs—captopril, methyldopa, and propranolol—on quality-of-life, long after their introduction into clinical practice.[7] Their report, which showed an advantage of captopril in several HRQL domains, had a major impact on drug prescription patterns at the time of its publication. The earlier in the process of drug development potential effects on quality-of-life are recognized, the sooner appropriate data may be collected and analyzed.

Methodologic problems to be addressed by pharmacoepidemiologic research

Researchers willing to accept the notion of the importance of measuring HRQL in pharmacoepidemiologic research and ready to use HRQL instruments in postmarketing (or, in some cases, premarketing) trials, face a considerable number of challenges. These challenges start with the realization that, as we have noted, there is no universal agreement on what the concept of quality-of-life actually entails. Thus, investigators must define as precisely as possible the aspects of HRQL in which they are interested.

Having identified the purpose for which an investigator wishes to use an HRQL instrument, one must be aware of the measurement properties required for it to fulfill its purpose. An additional problem occurs at this stage if researchers developed the original instrument in a different language, because one cannot assume the adequate performance of an instrument after its translation. At the next step, the investigator must choose from many available HRQL measurement instruments. When one has dealt satisfactorily with all these problems, the investigator has to ensure—as in any measurement—the rigorous fashion (standardized, reproducible, unbiased) with which to obtain the measurements (interviews or self- or computer-administered questionnaires). Finally, one is left with the chore of interpreting the data and translating the results into clinically meaningful terms.[8,9]

Currently available solutions

Quality-of-life measurement instruments in investigating new drugs: potential use and necessary attributes

In theory, any HRQL instrument could be used either to discriminate among patients (either according to current function or according to future prognosis), or to evaluate changes occurring in the health status (including HRQL) over time.[10,11] In most clinical trials, the primary objective of quality-of-life instruments is the evaluation of the effects of therapy, expressing treatment effects as a change in the score of the instrument over time. Occasionally, the intended use of instruments is to discriminate among patients. An example would be a study evaluating the effect of drug treatment on functional status in patients after myocardial infarction, where the investigators may wish to divide potential patients into those with moderate versus poor function (with a view toward intervening in the latter group).

The purpose for which investigators use an instrument dictates, to some degree, its necessary attributes. Each HRQL measurement instrument, regardless of its particular use, should be valid. The *validity* of an instrument refers to its ability to measure what it is intended to measure. This attribute of a measurement instrument is difficult to establish when there is no gold standard, as is the case with evaluation of HRQL. In such situations, where so-called *criterion validity* cannot be established, the validity of an instrument is frequently established in a step-wise process including examination of *face validity* (or sensibility)[12] and *construct validity*.

Face validity (sensibility) relies 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. Construct validity refers to the extent to which results from a given instrument relate to other measures in a manner consistent with theoretical hypotheses. It is useful to distinguish between *cross-sectional construct validity* and *longitudinal construct validity*. To explain the former one could hypothesize that scores on one HRQL instrument should correlate with scores on another HRQL instrument or a physiological measure when measured at one point in time. For example, for identification of patients with chronic airflow limitation who have moderate to severe functional status impairment, an instrument measuring patient-reported dyspnea should show correlation with spirometry. In contrast, one would anticipate that spirometry would discriminate less well between those with worse and better emotional function than it does between those with worse and better physical function. To exemplify longitudinal construct validity one could hypothesize that *changes* in spirometry related to a use of a new drug in patients with chronic airflow limitation should bear a close correlation with *changes* in functional status of the patient and a weaker correlation with *changes* in their emotional status.

The second attribute of an HRQL instrument is its ability to detect the "signal," over and above the "noise" which is introduced in the measurement process. For *discriminative instruments*, those that measure differences among people at a single point in time, this "signal" comes from differences between patients in HRQL. In this context, the way of quantifying the signal-to-noise ratio is called *reliability*. If the variability in scores between subjects (the signal) is much greater than the variability within stable subjects (the noise), an instrument will be deemed reliable. Reliable instruments will generally demonstrate that stable subjects show more or less the same results on repeated administration. The reliability coefficient (in general most appropriately an intraclass correlation coefficient) measuring the ratio of between-subject variance to total variance (which includes both between- and within-subject variance) is the statistic most frequently used to measure signal-to-noise ratio for discriminative instruments.

Classical reliability focuses on each observation or test score with a single true score, which belongs to one family of parallel observations, and yields a single reliability coefficient. Cronbach and colleagues introduced generalizability theory (G theory) as a framework for conceptualizing,

investigating, and designing reliable observations in response to limitations of the true-score-model of classical reliability theory.[13–15]

G theory acknowledges that in any measurement situation there are multiple, perhaps infinite, sources of error variance.[16] It involves the same assumptions as classical test theory, but is simply an extension that allows for a linear model including multiple sources of error. Application of G theory focuses on identifying and measuring these error variances. Once an investigator has identified all possible sources of error (e.g., domains or subcomponents of domains, raters, study centers), he or she can construct appropriate coefficients and show the extent to which one can generalize from observations made by one rater on one occasion about, for instance, a marker state (patient scenario), to the same rater (test–retest reliability) on a different occasion or to a different rater on the same occasion (inter-rater reliability). One assumption underlying G theory is that it is not possible to present a standard G theory design because each design may have different error sources. G theory provides an approach for dealing with these multiple sources of variance,[16] using multifactor repeated measures analysis of variance (ANOVA) to compute mean square terms for each main effect and interaction.[17]

For *evaluative instruments*, those designed to measure changes within individuals over time, the "signal" comes from the differences in HRQL within patients associated with the intervention. The way of determining the signal-to-noise ratio is called *responsiveness* and refers to an instrument's ability to detect change. If a treatment results in an important difference in HRQL, investigators may wish to be confident they will detect that difference, even if it is small. The responsiveness of an instrument is directly related to: (i) the magnitude of the difference in score in patients who have improved or deteriorated (the capacity to measure this signal can be called *changeability*), and (ii) the extent to which patients who have not changed obtain more or less the same scores (the capacity to minimize this noise can be called *reproducibility*). It follows that, to be of use, the ability of an instrument to show change when such change occurs has to be combined with its stability under unchanged conditions.

An example of an index of responsiveness is the ratio of the magnitude of change that corresponds to the minimally important difference (MID), to the variability in score in stable subjects.[18] Investigators have suggested other measurements of responsiveness, but they all rely on some way of relating signal to noise.[18–22]

Another essential measurement property of an instrument is the extent to which one can understand the magnitude of any differences between treatments that a study demonstrates—the instrument's *interpretability*. If a treatment improves HRQL score by 3 points relative to control, what are we to conclude? Is the treatment effect very large, warranting widespread dissemination in clinical practice, or is it trivial, suggesting the new treatment should be abandoned? This question highlights the importance of being able to interpret the results of HRQL questionnaire scores.

While our capacity to interpret results remains limited, investigators are adducing more and more information to enhance instrument interpretability.[8,9] Researchers have developed a number of strategies to address this difficult issue. Successful strategies have three things in common. First, they require an independent standard of comparison. Second, this independent standard must itself be interpretable. Third, there must be at least a moderate relationship between changes in questionnaire score and changes in the independent standard. The authors of this chapter have found that a correlation of 0.5 approximates the boundary between an acceptable and unacceptable relationship for establishing interpretability.

In our own work, we have often used global ratings of change (patients classifying themselves as unchanged, or experiencing small, medium, and large improvements, or deteriorations) as the independent standard. We construct our disease-specific instruments using 7-point scales with an associated verbal descriptor for each level on the scale. For each questionnaire domain, we divide the total score by the number of items so that domain scores can range from 1 to 7. Using this approach to framing response options, we have found that the

smallest difference that patients consider important is often approximately 0.5 per question.[21,23] A moderate difference corresponds to a change of approximately 1.0 per question, and changes of greater than 1.5 can be considered large. So, for example, in a domain with four items, patients will consider a one point change in two or more items as important. This finding seems to apply across different areas of function, including dyspnea, fatigue, and emotional function in patients with chronic airflow limitation;[21] symptoms, emotional function, and activity limitations in both adult[23] and child[24] asthma patients, and parents of child asthma patients;[25] and symptoms, emotional function, and activity limitations in adults with rhinoconjunctivitis.[26] Similar observations may be derived from reports of other investigators.[27]

The approach that we have just described relies on within-patient comparisons as the independent standard. An alternative is between-patient comparisons. In one example of this approach, we formed groups of seven patients with chronic airflow limitation participating in a respiratory rehabilitation program.[28] Each patient completed the Chronic Respiratory Questionnaire (CRQ). The patients conversed with one another long enough to make judgments about their relative experience of fatigue in daily life. While there was a bias in their assessment (patients generally considered themselves better off than one another), their relative ratings allow estimates of what differences in CRQ score constitute small, medium, and large differences. The results were largely congruent with the findings from the within-patient rating studies.[28]

Another anchor-based approach uses HRQL instruments for which investigators have established the minimal important difference (MID). Investigators can apply regression or other statistical methods to compute the changes on a new instrument that correspond to those of the instrument with the established MID. For example, using the established MID of the CRQ we computed the MID for two other instruments that measure HRQL in patients with chronic airflow limitation, the feeling thermometer and the St George's Respiratory Questionnaire.[29] Similar to the anchor-based approach using transition ratings, investigators should ensure that the strength of the correlation between the change scores of these instruments exceeds a minimum (for example, a correlation coefficient of 0.5).

Yet another approach to estimate the MID involves enrolling panels of experts or patients and using qualitative research methods, such as Delphi techniques. Using a panel-based approach, Wyrwich *et al.* enrolled pulmonary physicians to determine the MID of the CRQ.[30] All panel members were familiar with the CRQ, received information about the instrument, and received materials about the previously established MID for the instrument. The experts came to a consensus on what constitutes the MID of the CRQ. The results for the MID were similar to those obtained with the anchor-based approach described above (a change of 0.5 on the 7-point scale).

Investigators have proposed distribution-based methods to determine interpretability of HRQL instruments. Distribution-based methods differ from anchor-based methods in that they interpret results in terms of the relation between the magnitude of effect and some measure or measures of variability in results.[9] The magnitude of effect can be the difference in an individual patient's score before and after treatment, a single group's score before and after treatment, or the difference in score between treatment and control groups. As a measure of variability, investigators may choose between-patient variability (the standard deviation of patients at baseline, for instance) or within-patient variability (the standard deviation of change that patients experienced during a study).

If an investigator used the distribution-based approach, the clinician would see a treatment effect reported as, for instance, 0.3 standard deviation units. The great advantage of distribution-based methods is that the values are easy to generate for almost any HRQL instrument because there will always be one or more measures of variability available. This contrasts with the work needed to generate an anchor-based interpretation, evident from the prior discussion. The problem related to this methodology is that the units do not have intuitive meaning to clinicians. It is possible, however, that clinicians could gain experience with

standard deviation units in the same way they learn to understand other HRQL scores.

Cohen addressed this problem in a seminal work by suggesting that changes in the range of 0.2 standard deviation units represent small changes, those in the range of 0.5 standard deviation units represent moderate changes, and those in the range of 0.8 standard deviation units represent large changes.[31] Thus, one would tell a clinician that if trial results show a 0.3 standard deviation difference between treatment and control, then the patient can anticipate a small improvement in HRQL with treatment. The problem with this approach is the arbitrariness. Do 0.2, 0.5, and 0.8 standard deviation units consistently represent small, medium, and large effects?

In response to this problem, investigators have attempted to provide empirical evidence about the relationship between distribution-based and anchor-based results. These studies address the question, "What is the appropriate interpretation of a particular magnitude of effect in distribution-based units, as judged by the results of anchor-based studies?" For example, we described the MID for the CRQ based on Cohen's effects size in patients completing a respiratory rehabilitation program.[29] The CRQ scores that corresponded to 0.2, 0.5, and 0.8 standard deviation units were as follows: CRQ dyspnea 0.24, 0.61, and 0.98; CRQ fatigue 0.27, 0.67, and 1.08; CRQ emotional function 0.24, 0.60, and 0.96; and CRQ mastery 0.24, 0.60, and 0.96. Thus, this work indicates the MID to be in the range of 0.2 to 0.5 standard deviation units.

The standard error of measurement (SEM) presents another distribution-based method. It is defined as the variability between an individual's observed score and the true score, and is computed as the baseline standard deviation multiplied by the square root of 1 minus the reliability of the QOL measure. In theory, a QOL measure's standard error of measurement is sample independent, whereas its component statistics, the standard deviation and the reliability estimate, are sample dependent and vary around the standard error of measurement.[32] When the between-person variability in the population increases, the standard deviation will increase (tending to raise the standard error of measurement), but the reliability will also increase (tending to lower the standard error of measurement). Thus, the standard error of measurement largely reflects within-person variability over time. Wyrwich and colleagues provide an example of using the SEM approach in their study following 471 outpatients with chronic obstructive pulmonary disease. The authors used the SEM to correlate this distribution-based method with the MID.[33] They found that the SEM method consistently suggested an MID of the CRQ of approximately 0.5. In addition, the research revealed that this methodology shows consistent estimates for the MID across a wide range of HRQL scores on the CRQ.

Using four prospectively collected longitudinal data sets for four established HRQL instruments (i.e., Pediatric Ulcerative Colitis Activity Index, Pediatric Crohn's Disease Activity Index, the Rhinoconjunctivitis Quality of Life Questionnaire, and the Chronic Respiratory Questionnaire), a recent study compared anchor-based methods versus five distribution-based methods for establishing MID.[34] For HRQL instruments, 0.5 and 1 SEM seemed to provide values closest to the anchor-based estimates of MID. However, the investigators found a lack of consistency between anchor and distribution-based methods and substantial variability between proposed distribution-based approaches. Given this lack of consistency, authors argue that distribution-based approaches should act only as a provisional substitute, pending availability of empirically established anchor-based estimates.[34]

Investigators can also use as independent standards measures that clinicians, through long experience, already know well. For example, scores on a generic measure of HRQL, the Sickness Impact Profile (SIP), range from an average of 8.2 in patients with American Rheumatism Association arthritis class I, to 25.8 in class IV.[35] Another standard would be obtained by administering questionnaires to patients before and after an intervention of known effectiveness with which clinicians are familiar, so that they can see the change in score associated with response to treatment. For example, patients shortly after hip replacement have scores

of 30 on the SIP, scores which decrease to less than 5 after full convalescence.[36] Relationships between HRQL and a variety of marker states can also be useful: SIP scores in patients with chronic airflow limitation severe enough to require home oxygen are approximately 24;[37] scores in patients with chronic, stable angina are approximately 11.5.[38]

Clinicians and investigators tend to assume that, if the mean difference between a treatment and a control is appreciably less than the smallest change that is important, then the treatment has a trivial effect. This may not be so. Let us assume that a randomized clinical trial (RCT) shows a mean difference of 0.25 in a questionnaire with an MID of 0.5. One may conclude that the difference is unimportant, and the result does not support administration of the treatment. This interpretation assumes that every patient given treatment scored 0.25 better than they would have, had they received the control. However, it ignores possible heterogeneity of the treatment effect. Depending on the true distribution of results, the appropriate interpretation may be different.

Consider a situation where 25% of the treated patients improved by a magnitude of 1.0, while the other 75% did not improve at all (mean change of 0). This would indicate that the 25% of treated patients obtained moderate benefit from the intervention. Using the number needed to treat (NNT), a methodology developed for interpreting the magnitude of treatment effects, investigators have found that clinicians commonly treat 25 to 50 patients, and often as many as 100, to prevent a single adverse event.[39] Thus, the hypothetical treatment with a mean difference of 0.25 and an NNT of 4 proves to have a powerful effect.

We have shown that this issue is much more than hypothetical.[40] In a crossover randomized trial in asthmatic patients comparing the short-acting inhaled beta-agonist salbutamol to the long-acting inhaled beta-agonist salmeterol, we found a mean difference of 0.3 between groups in the activity dimension of the Asthma Quality-of-Life Questionnaire (AQLQ). This mean difference represents slightly more than half the minimal important difference in an individual patient. Knowing that the minimal important difference is 0.5 allows us to calculate the proportion of patients who achieved benefit from salmeterol—that is, the proportion who had an important improvement (greater than 0.5 in one of the HRQL domains) while receiving salmeterol relative to salbutamol. For the activity domain of the AQLQ, this proportion proved to be 0.22 (22%). The NNT is simply the inverse of the proportion who benefit, in this case 4.5. Thus, clinicians need to treat fewer than five patients with salmeterol to ensure that one patient obtains an important improvement in their ability to undertake activities of daily living.

In another randomized trial examining the effect of a respiratory rehabilitation program in patients with chronic lung disease, we found a mean difference between rehabilitation patients and the community controls of 0.40 in the emotions domain of the CRQ.[41] This difference is less than the value of 0.5 that represents the minimal important difference in an individual patient. However, the data from the trial allowed us to calculate the proportion of patients who were ≥0.5 better in their emotional function while receiving rehabilitation than would have been the case had they been in the community control group. This turned out to be 0.30, which translates into an NNT of 4 (or exactly 3.3) patients.

This discussion emphasizes that, to interpret the results of HRQL measurement in pharmacoepidemiologic studies requires clinicians to be aware of the changes in score that constitute trivial, small, medium, and large differences in HRQL. Further, looking at mean differences between groups can be misleading. The distribution of differences is critical, and can be summarized in an informative manner using the NNT.

Quality-of-life measurement instruments: taxonomy and potential use

Clinical journals have published trials in which HRQL instruments are the primary outcome measures. With the expanding importance of HRQL in evaluating new therapeutic interventions, investigators (and readers) are faced with a large array of instruments. Researchers have proposed different ways of categorizing these instruments, according

to the purpose of their use; into instruments designed for screening, providing health profiles, measuring preference, and making clinical decisions;[42] or into discriminative and evaluative instruments (as above).

We have also suggested a taxonomy based on the domains of HRQL which an instrument attempts to cover.[43] According to this taxonomy, an HRQL instrument may be categorized, in a broad sense, as generic or specific. *Generic instruments* cover (or at least aim to cover) the complete spectrum of function, disability, and distress of the patient, and are applicable to a variety of populations and conditions. Within the framework of generic instruments, health profiles and utility measures provide two distinct approaches to measurement of global quality-of-life. *Specific instruments* are focused on disease or treatment issues particularly relevant to the disease or condition of interest.

Generic instruments

Health profiles

Health profiles are single instruments that measure multiple different aspects of quality-of-life. They usually provide a scoring system that allows aggregation of the results into a small number of scores and sometimes into a single score (in which case, it may be referred to as an index). As generic measures, they are designed for use in a wide variety of conditions. For example, one health profile, the SIP contains 12 "categories" which can be aggregated into two dimensions and five independent categories, and also into a single overall score.[36] The SIP has been used in studies of cardiac rehabilitation,[44] total hip joint arthroplasty,[45] and treatment of back pain.[46] In addition to the SIP, there are a number of other health profiles available: the Nottingham Health Profile,[47] the Duke–UNC Health Profile,[48] and the McMaster Health Index Questionnaire.[49] Increasingly, a collection of related instruments from the Medical Outcomes Study[50] have become the most popular and widely used generic instruments. Particularly popular is one version that includes 36 items, the SF-36.[51–53] The SF-36 is available in over 40 languages, and normal values for the general population in many countries are available.

While each health profile attempts to measure all important aspects of HRQL, they may slice the HRQL pie quite differently. For example, the McMaster Health Index Questionnaire follows the World Health Organization approach and identifies three dimensions: physical, emotional, and social. The SIP includes a physical dimension (with categories of ambulation, mobility, body care, and movement), a psychosocial dimension (with categories including social interaction and emotional behavior), and five independent categories including eating, work, home management, sleep and rest, and recreations and pastimes.

General health profiles offer a number of advantages to the clinical investigator. Their reproducibility and validity have been established, often in a variety of populations. When using them for discriminative purposes, one can examine and establish areas of dysfunction affecting a particular population. Identification of these areas of dysfunction may guide investigators who are constructing disease-specific instruments to potentially target areas with the greatest impact on the quality-of-life. Health profiles, used as evaluative instruments, allow determination of the effects of an intervention on different aspects of quality-of-life, without necessitating the use of multiple instruments (which may save both the investigator's and the patient's time). Because health profiles are designed for a wide variety of conditions, one can potentially compare the effects on HRQL of different interventions in different diseases. Profiles that allow computation of a single score can be used in a cost-effectiveness analysis, in which the cost of an intervention in dollars is related to its outcome in natural units (see Chapter 38).

The main limitation of health profiles is that they may not focus adequately on the aspects of quality-of-life specifically influenced by a particular intervention. This may result in an inability of the instrument to detect a real effect in the area of importance (i.e., lack of responsiveness). In fact, disease-specific instruments offer greater responsiveness compared with generic instruments.[54,55] We will return to this issue when we discuss the alternative approach, specific instruments.

Utility measurement

Economic and decision theory provides the underlying basis for *utility measures* (see Chapter 38). The key elements of a utility instrument are, first, that it is preference-based (i.e., based on patients' personal preferences) and, second, that scores are tied to death as an outcome. Typically, HRQL can be measured as a utility measure using a single number along a continuum from dead (0.0) to full health (1.0). The use of utility measures in clinical studies requires serial measurement of the utility of the patient's quality-of-life throughout the study.

There are two fundamental approaches to utility measurement in clinical studies. One is to ask patients a number of questions about their function and well-being. Based on their responses, patients are classified into one of a number of categories. Each category has a utility value associated with it, the utility having been established in previous ratings by another group (ideally a random sample of the general population). This approach is typified by three widely used instruments: the Quality of Well-Being Scale,[56–58] the Health Utilities Index,[59] and the Euroqol (EQ5).[60]

The second approach is to ask patients to make a single rating that takes into account all aspects of their quality-of-life.[61] This rating can be made in many ways. The "standard gamble" asks patients to choose between their own health state and a gamble in which they may die immediately or achieve full health for the remainder of their lives. Using the standard gamble, patients' utility or HRQL is determined by the choices they make, as the probabilities of immediate death or full health are varied. Another technique is the "time trade-off," in which subjects are asked about the number of years in their present health state they would be willing to trade-off for a shorter life span in full health. A third technique is the use of a simple visual analogue scale presented as a thermometer, the "feeling thermometer."[62] When completing the feeling thermometer, patients choose the score on the thermometer that represents the value they place on their health state. The best state is full health (equal to a score of 100) and the worst state is dead (a score of 0).

A major advantage of utility measurement is its amenability to cost–utility analysis (see Chapter 38). In cost–utility analysis, the cost of an intervention is related to the number of quality-adjusted life-years (QALYs) gained through application of the intervention.[63] Cost per QALY may be compared and provides a basis for allocation of scarce resources among different health-care programs. Results from the utility approach may thus be of particular interest to program evaluators and health policy decision makers.

However, utility measurement also has limitations. Utilities can vary depending on how they are obtained, raising questions of the validity of any single measurement.[64,65] Utility measurement does not allow the investigator to determine which aspects of HRQL are responsible for changes in utility. Finally, utilities potentially share the disadvantage of health profiles, in that they may not be responsive to small but still clinically important changes.

Specific instruments

An alternative approach to HRQL measurement is to focus on aspects of health status that are specific to the area of primary interest. The rationale for this approach lies in the increased responsiveness that may result from including only those aspects of HRQL that are relevant and important in a particular disease process or even in a particular patient situation. One could also focus an instrument only on the areas that are likely to be affected by a particular drug. This latter approach is advanced in the design and conduct of randomized controlled trials with individual patients—*N*-of-1 randomized clinical trials.[66]

In other situations, the instrument may be specific to the disease (e.g., for chronic lung disease, for rheumatoid arthritis, for cardiovascular diseases, for endocrine problems); specific to a population of patients (e.g., the frail elderly, who are afflicted with a wide variety of different diseases); specific to a certain function (e.g., emotional or sexual function); or specific to a given condition or problem (e.g., pain), which can be caused by a variety of underlying pathologies. Within a single

condition, the instrument may differ depending on the intervention. For example, while success of a disease modifying agent in rheumatoid arthritis should result in improved HRQL by enabling a patient to increase performance of physically stressful activities of daily living, occupational therapy may achieve improved HRQL by encouraging family members to take over activities formerly accomplished with difficulty by the patient. Appropriate disease-specific HRQL outcome measures should reflect this difference.

Specific instruments can be constructed to reflect the "single state" ("How tired have you been: very tired, somewhat tired, full of energy?") or a "transition" ("How has your tiredness been: better, the same, worse?").[67] Theoretically, the same could be said of generic instruments, although none of the available generic instruments has used the transition approach. Specific measures can integrate aspects of morbidity, including events such as recurrent myocardial infarction.[68]

Like generic instruments, disease-specific instruments may be used for discriminative purposes. They may aid, for example, in evaluating the extent to which a primary symptom (for example dyspnea) is related to the magnitude of physiological abnormality (for example exercise capacity).[69] Disease-specific instruments can be applied for evaluative purposes to establish the impact of an intervention on a specific area of dysfunction, and hence aid in elucidating the mechanisms of drug action.[70] Guidelines provide structured approaches for constructing specific measures.[71] Whatever approaches one takes to the construction of disease-specific measures, a number of head-to-head comparisons between generic and specific instruments suggest that the latter approach will fulfill its promise of enhancing responsiveness.[27,72–76]

In addition to the improved responsiveness, specific measures have the advantage of relating closely to areas routinely explored by the physician. For example, a disease-specific measure of quality-of-life in chronic lung disease focuses on dyspnea during day-to-day activities, fatigue, and areas of emotional dysfunction, including frustration and impatience.[18] Specific measures may therefore appear clinically sensible to the clinician.

The disadvantages of specific measures are that they are (deliberately) not comprehensive, and cannot be used to compare across conditions or, at times, even across programs. This suggests that there is no one group of instruments that will achieve all the potential goals of HRQL measurement. Thus, investigators may choose to use multiple instruments, an issue we will deal with in the next section.

Use of multiple quality-of-life measures in clinical studies

Clinical investigators are not restricted to using a single instrument in their studies, and investigators will often conclude that a single instrument cannot yield all the relevant information. For example, utility and disease-specific measures contribute quite different sorts of data, and an investigator may want to use one of each.

Another, somewhat different way of using multiple instruments is to administer a battery of specific instruments. An example of such an approach was a blinded, randomized trial of three antihypertensive agents in primary hypertension.[7] The investigators identified five dimensions of health they were measuring: the sense of well-being and satisfaction with life, the physical state, the emotional state, intellectual functioning, and the ability to perform in social roles and the degree of satisfaction from those roles. Even within these five dimensions, additional components were present. For example, separate measurements of sleep and sexual function existed. Patients taking one of the three drugs under investigation, captopril, scored better on measures of general well-being, work performance, and life satisfaction. The lesson for the clinician is clearly important: one can have an impact on not only the length, but also the quality of the patient's life according to choice of antihypertensive agent.

The approach of using multiple instruments, although comprehensive, has limitations. First, investigators must find a valid, responsive instrument for every attribute they wish to measure. Second, it is possible (indeed likely) that only some of the instruments chosen will show differences between the treatments under investigation. Unless

one of the instruments has been designated as the primary measure of outcome before the study started, different results in different measures may make interpretation difficult. The greater the number of instruments used, the greater the probability that one or more instruments will favor one treatment or the other, even if the treatments' true effectiveness is identical. Thus, the alpha error (the probability of finding an apparent difference between treatments when in fact their outcomes do not differ) increases with each new instrument used. Although this problem may be dealt with through statistical adjustment for the number of instruments used, such adjustment is often not made.[77]

Another problem occurs if only a small proportion of the instruments used favor an intervention (or if some instruments favor one treatment and other instruments favor the other). In these situations, the clinician may be unsure how to interpret the results. The use of multiple instruments opens the door to such potential controversy.

A final limitation of using a battery of instruments is that it gives no indication of the relative importance of various areas of dysfunction to the patient. For example, had Croog *et al.*[7] found that one antihypertensive agent disturbed sleep, while another had an adverse impact on sexual function, their approach would not have allowed determination of which drug had a greater net adverse impact on patients' lives.

The future

The considerations we have raised suggest a step-by-step approach to addressing issues of HRQL in pharmacoepidemiologic studies.[78–81] Clinicians must begin by asking themselves if investigators have addressed all the important effects of treatment on patients' quantity and quality-of-life. If they have not, clinicians may have more difficulty applying the results to their patients.

If the study has addressed HRQL issues, have investigators chosen the appropriate instruments? In particular, does evidence suggest the measures used are valid measures of HRQL? If so, and the study failed to demonstrate differences between groups, is there good reason to believe the instrument is responsive in this context? If not, the results may be a false negative, failing to show the true underlying difference in HRQL.

Whatever the differences between groups, the clinician must be able to interpret their magnitude. Knowledge of the difference in score that represents small, medium, and large differences in HRQL will be very helpful in making this interpretation. Clinicians must still look beyond mean differences between groups, and consider the distribution of differences. The NNT for a single patient to achieve an important benefit in HRQL offers one way of expressing results that clinicians are likely to find meaningful.

References

1. Franciosa J, Jordan R, Wilen M. Minoxidil in patients with chronic left heart failure: contrasting hemodynamic and clinical effects in a controlled trial. *Circulation* 1984; **70**: 63–8.
2. Packer M. How should we judge the efficacy of drug therapy in patients with chronic congestive heart failure? The insight of the six blind men. *J Am Coll Cardiol* 1987; **9**: 433–8.
3. Danis M, Gerrity M, Southerland L, Patrick D. A comparison of patient, family, and physician assessment of the value of medical intensive care. *Crit Care Med* 1988; **16**: 594–600.
4. Schipper H, Clinch J, Powell V. Definitions and conceptual issues. In: Spilker B, ed. *Quality-of-Life Assessment in Clinical Trials*. New York: Raven Press, 1990, pp. 11–24.
5. Gill T, Feinstein A. A critical appraisal of the quality of quality-of-life measurements. *JAMA* 1994; **272**: 619–26.
6. Guyatt G, Cook D. Health status, quality-of-life, and the individual patient. A commentary on: Gill TM, Feinstein AR. A critical appraisal of quality-of-life measurements. *JAMA* 1994; **272**: 630–1.
7. Croog S, Levine S, Testa M. The effects of antihypertensive therapy on the quality-of-life. *N Engl J Med* 1986; **314**: 1657–64.
8. Guyatt GH. Making sense of quality-of-life data. *Med Care* 2000; **38** (Suppl II): 175–9.

9. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting G. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; **77**: 371–83.
10. Kirshner B, Guyatt G. A methodologic framework for assessing health care indices. *J Chronic Dis* 1985; **38**: 27–36.
11. Guyatt G, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992; **45**: 1341–5.
12. Feinstein A. The theory and evaluation of sensibility. In: Feinstein A, ed. *Clinimetrics*. New Haven, CT: Yale University Press, 1987, pp. 141–65.
13. Cronbach LJ, Nageswari R, Gleser GC. Theory of generalizability: a liberation of reliability theory. *Br J Stat Psychol* 1963; **16**: 137–63.
14. Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley, 1972.
15. Spearman C. The proof and measurement of association between two things. *Am J Psychol* 1904; **15**: 72–101.
16. Streiner DL, Norman GR. Generalizability theory. In: Streiner DL, Norman GR, eds. *Health Measurement Scales*. Oxford: Oxford University Press, 2008, pp. 211–46.
17. Schünemann HJ, Norman G, Puhan MA, Stahl E, Griffith L, Heels-Ansdell D, *et al.* Application of generalizability theory confirmed lower reliability of the standard gamble than the feeling thermometer. *J Clin Epidemiol* 2007; **60**: 1256–62.
18. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987; **40**: 171–8.
19. Levine M. Incorporation of quality-of-life assessment into clinical trials. In: Osoba D, ed. *Effect of Cancer on Quality-of-Life*. Boston, MA: CRC Press, 1991, pp. 105–11.
20. Guyatt G, Feeny D, Patrick D. Measuring health-related quality-of-life: basic sciences review. *Ann Intern Med* 1993; **70**: 225–30.
21. Jaeschke R, Guyatt G, Singer J. Measurement of health status: ascertaining the meaning of a change in quality-of-life questionnaire score. *Control Clin Trials* 1989; **10**: 407–15.
22. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000; **53**: 459–68.
23. Juniper E, Guyatt G, Willan A, Griffth L. Determining a minimal important change in a disease-specific quality-of- life questionnaire. *J Clin Epidemiol* 1994; **47**: 81–7.
24. Juniper E, Guyatt G, Feeny D, Ferrie PJ, Griffith LE, Townsend M. Measuring quality-of-life in children with asthma. *Qual Life Res* 1996; **5**: 35–46.
25. Juniper E, Guyatt G, Feeny D, Ferrie P, Griffith L, Townsend M. Measuring quality-of-life in the parents of children with asthma. *Qual Life Res* 1996; **5**: 27–34.
26. Juniper EF, Guyatt GH, Griffith LE, Ferrie PJ. Interpretation of rhinoconjunctivitis quality-of-life questionnaire data. *J Allergy Clin Immunol* 1996; **98**: 843–5.
27. Laupacis A, Wong C, Churchill D. The use of generic and specific quality-of-life measures in hemodialysis patients treated with erythropoietin. *Control Clin Trials* 1991; **12**: 168S–79S.
28. Redelmeier D, Guyatt G, Goldstein R. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996; **49**: 1215–19.
29. Schunemann H, Griffith L, Jaeschke R, Stubbing D, Goldstein R, Guyatt GH. Evaluation of the minimal important difference for the feeling thermometer and St. George's Respiratory Questionnaire in patients with chronic airflow limitation. *J Clin Epidemiol* 2003; **56**: 1170–6.
30. Wyrwich KW, Fihn SD, Tierney WM, Kroenke K, Babu AN, Wolinsky FD. Clinically important changes in health- related quality-of-life for patients with chronic obstructive pulmonary disease. An expert consensus panel report. *J Gen Intern Med* 2003; **18**: 196–202.
31. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
32. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality-of-life. *Med Care* 1999; **37**: 469–78.
33. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality-of-life. *J Clin Epidemiol* 1999; **52**: 861–73.
34. Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, *et al.* The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol* 2010; **63**: 28–36.

35. Deyo R, Inui T, Leininger J, Overman S. Measuring functional outcomes in chronic disease: a comparison of traditional scales and a self-administered health status questionnaire in patients with rheumatoid arthritis. *Med Care* 1983; **21**: 180–92.
36. Bergner M, Bobbitt R, Carter W, Gilson B. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981; **19**: 787–805.
37. McSweeney A, Grant I, Heaton R, Adams KM, Timms RM. Life quality of patients with chronic obstructive pulmonary disease. *Arch Intern Med* 1982; **142**: 473–8.
38. Fletcher A, McLoone P, Bulpitt C. quality-of-life on angina therapy: a randomised controlled trial of transdermal glyceryl trinitrate against placebo. *Lancet* 1988; **2**: 4–7.
39. Laupacis A, Sackett D, Roberts R. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; **318**: 1728–33.
40. Guyatt G, Juniper E, Walter S, Griffith L, Goldstein R. Interpreting treatment effects in randomized trials. *BMJ* 1998; **316**: 690–3.
41. Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW. A measure of quality-of-life for clinical trials in chronic lung disease. *Thorax* 1987; **42**: 773–8.
42. Osoba D. *Effect of Cancer on Quality-of-Life*. Boston, MA: CRC Press, 1991.
43. Guyatt G, Zanten SVV, Feeny D, Patrick D. Measuring quality-of-life in clinical trials: a taxonomy and review. *Can Med Assoc J* 1989; **140**: 1441–7.
44. Ott C, Sivarajan E, Newton K, Almes MJ, Bruce RA, Bergner M, *et al.* A controlled randomized study of early cardiac rehabilitation: the sickness impact profile as an assessment tool. *Heart Lung* 1983; **12**: 162–70.
45. Liang M, Larson M, Cullen K, Schwartz J. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985; **28**: 542–7.
46. Deyo R, Diehl A, Rosenthal M. How many days of bed rest for acute low back pain? A randomized clinical trial. *N Engl J Med* 1986; **315**: 1064–70.
47. Hunt S, McKenna S, McEwen J, Backett E, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health* 1980; **34**: 281–6.
48. Parkerson G, Gehlback S, Wagner E, Sherman A, Clapp N, Muhlbaier L. The Duke–UNC Health Profile: an adult health status instrument for primary care. *Med Care* 1981; **19**: 806–28.
49. Sackett D, Chambers L, MacPherson A, Goldsmith C, McAuley R. The development and application of indices of health: general methods and a summary of results. *Am J Public Health* 1977; **67**: 423–8.
50. Tarlov A, Ware J, Greenfield S, Nelson E, Perrin E, Zubkoff M. The Medical Outcomes Study. *JAMA* 1989; **262**: 925–30.
51. Brook R, Ware J, Davies-Avery A, Stewart A, Donald C, Rogers WH, *et al.* Overview of adult health status measures fielded in Rand's health insurance study. *Med Care* 1979; **17** (Suppl 7): 1–131.
52. Ware J, Sherbourne C. The MOS 36-item short-form health survey (SF-36). *Med Care* 1992; **30**: 473–83.
53. Ware J, Kozinski M, Bayliss M, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results of the Medical Outcomes Study. *Med Care* 1995; **33**: AS264–79.
54. Guyatt G, King DR, Feeny DH, Stubbing D, Goldstein RS. Generic and specific measurement of health-related quality-of-life in a clinical trial of respiratory rehabilitation. *J Clin Epidemiol* 1999; **52**: 187–92.
55. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of- life instruments. *J Clin Epidemiol* 2003; **56**: 52–60.
56. Kaplan R, Anderson J, Wu A, Mathews W, Kozin F, Orenstein D. The quality of well-being scale: applications in AIDS, cystic fibrosis, and arthritis. *Med Care* 1989; **27** (Suppl): S27–43.
57. Kaplan R, Bush J, Berry C. Health status: types of validity and the index of well-being. *Health Serv Res* 1976; **11**: 478–507.
58. Kaplan R, Bush J. Health-related quality-of-life measurement for evaluation research and policy analysis. *Health Psychol* 1982; **1**: 61–80.
59. Feeny D, Furlong W, Boyle M, Torrance G. Multi-attribute health status classification systems: health utilities index. *Pharmacoeconomics* 1995; **7**: 490–502.
60. Anonymous. EuroQol—a new facility for the measurement of health-related quality-of-life. The EuroQol Group. *Health Policy* 1990; **16**: 199–208.
61. Torrance G. Measurement of health state utilities for economic appraisal: a review. *J Health Econ* 1986; **5**: 1–30.
62. Bennett K. Measuring health state preferences and utilities: rating scale, time trade-off, and standard gamble techniques. In: Spilker B, ed. *Quality-of-life and Pharmacoeconomics in Clinical Trials*. Philadelphia, PA: Lippincott-Raven, 1996, p. 259.

63. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making* 1989; **9**: 142–9.
64. Sutherland H, Dunn V, Boyd N. Measurement of values for states of health with linear analog scales. *Med Decis Making* 1983; **3**: 477–87.
65. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. The measurement of patients' values in medicine. *Med Decis Making* 1982; **3**: 477–87.
66. Guyatt G, Keller J, Jaeschke R, Rosenbloom D, Adachi J, Newhouse M. The *N*-of-1 randomized controlled trials: clinical usefulness. Three year experience. *Ann Intern Med* 1990; **112**: 293–9.
67. MacKenzie M, Charlson M. Standards for the use of ordinal scales in clinical trials. *BMJ* 1986; **292**: 40–3.
68. Olsson G, Lubsen J, Van Es G, Rehnqvist N. Quality-of-life after myocardial infarction: effect of long term metoprolol on mortality and morbidity. *BMJ* 1986; **292**: 1491–3.
69. Mahler D, Rosiello R, Harver A, Lentine T, McGovern J, Daubenspeck J. Comparison of clinical dyspnea ratings and psychophysical measurements of respiratory sensation in obstructive airway disease. *Am Rev Respir Dis* 1987; **135**: 1229–33.
70. Jaeschke R, Singer J, Guyatt G. Using quality-of-life measures to elucidate mechanism of action. *Can Med Assoc J* 1991; **144**: 35–9.
71. Guyatt G, Bombardier C, Tugwell P. Measuring disease-specific quality-of-life in clinical trials. *Can Med Assoc J* 1986; **134**: 889–95.
72. Tandon P, Stander H, Schwarz RP Jr. Analysis of quality-of-life data from a randomized, placebo controlled heart- failure trial. *J Clin Epidemiol* 1989; **42**: 955–62.
73. Smith D, Baker G, Davies G, Dewey M, Chadwick. D. Outcomes of add-on treatment with Lamotrigine in partial epilepsy. *Epilepsia* 1993; **34**: 312–22.
74. Chang S, Fine R, Siegel D, Chesney M, Black D, Hulley S. The impact of diuretic therapy on reported sexual function. *Arch Intern Med* 1991; **151**: 2402–8.
75. Tugwell P, Bombardier C, Buchanan W, Goldsmith C, Grace E, Bennett K, *et al.* Methotrexate in rheumatoid arthritis. Impact on quality-of-life assessed by traditional standard-item and individualized patient preference health status questionnaires. *Arch Intern Med* 1990; **150**: 59–62.
76. Goldstein R, Gort E, Guyatt G, Stubbing D, Avendano M. Prospective randomized controlled trial of respiratory rehabilitation. *Lancet* 1994; **344**: 1394–7.
77. Pocock S, Hughes M, Lee R. Statistical problems in the reporting of clinical trials. *N Engl J Med* 1987; **317**: 426–32.
78. Furukawa TA, Jaeschke R, Cook DJ, Guyatt G. Advanced topics in results of therapy trials: Measuring patients experience. In: Guyatt G, Rennie D, Meade MO, Cook DJ, eds. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice*. New York, NY: McGraw-Hill Medical, 2008, pp. 249–71.
79. Guyatt G. Therapy and harm: an introduction. In: Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago, IL: American Medical Association Press, 2002, pp. 49–100.
80. Guyatt G, Naylor C, Juniper E, Keyland D, Jaeschke R, Cook D. Users' guides to the medical literature: XII. How to use articles about health-related quality-of-life. *JAMA* 1997; **277**: 1232–7.
81. Dans A, Dans L, Guyatt G, Richardson S. For the Evidence-based Medicine Working Group. User's guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. *JAMA* 1998; **279**: 545–9.

CHAPTER 40
The Use of Meta-analysis in Pharmacoepidemiology

Jesse A. Berlin,[1] *M. Soledad Cepeda,*[1] *and Carin J. Kim*[2]
[1] Department of Epidemiology, Johnson & Johnson Pharmaceutical Research and Development, Titusville, NJ, USA
[2] Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA

Introduction

Definitions

Meta-analysis 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] 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.[3] While epidemiologists have been cautious about adopting meta-analysis, because of the inherent biases in the component studies and the great diversity in study designs and populations,[4–6] 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.[6–11] Meta-analysis of randomized clinical trials has found such wide acceptance that an entire international organization, the Cochrane Collaboration, has been built around the performance and updating of systematic reviews and meta-analyses of trials.[12] Cochrane Reviews are maintained in a publicly available electronic library. More information is available on the Cochrane web site (http://www.cochrane.org). A similar structure has developed in the social sciences, in the form of the Campbell Collaboration.[13]

Meta-analysis may be regarded as a "state-of-the-art" literature review, employing statistical methods in conjunction with a thorough and systematic qualitative review.[14] The distinguishing feature of meta-analysis, as opposed to the usual qualitative literature review, is its systematic, structured, and presumably objective presentation and analysis of available data. The traditional review has been increasingly recognized as being subjective.[14–17] With the support of leading scientists[18] and journal editors,[19] there has been growing acceptance of the concept that the literature review can be approached as a more rigorous scientific endeavor, specifically, an observational study with the same requirements for planning, prespecification of definitions, use of eligibility definitions, etc., as any other observational study. In recent years, the terms "research synthesis" and "systematic review" have been used to describe the structured review process in general, while "meta-analysis" has been reserved for the quantitative aspects of the process. For the purposes of this chapter, we shall use "meta-analysis" in the more general sense. Meta-analysis provides the conceptual and quantitative framework for such rigorous literature reviews; similar measures from comparable studies are tabulated systematically and the effect measures are combined when appropriate.

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

Several activities may be included under the above definition of meta-analysis. Perhaps the most popular conception of meta-analysis, 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 meta-analysis that examined the effects of aspirin following myocardial infarction. Typically, this type of meta-analysis would present an overall measure of the efficacy of treatment, for example a summary odds ratio. Summary measures may be presented for different subsets of trials involving specific types of patients, for example studies restricted to men versus studies that include both men and women. More sophisticated meta-analyses also examine the variability of results among trials and, when results have been conflicting, attempt to uncover the sources of the disagreements.[20]

More recently, meta-analyses of non-experimental epidemiologic studies have been performed,[21–24] and articles have been written describing the methodologic considerations specific to those meta-analyses.[25–32] In general, both the meta-analyses of non-experimental 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 of non-experimental studies, it is logical to find more disagreement among non-experimental studies than among randomized trials.

This chapter summarizes many of the major conceptual and methodologic issues surrounding meta-analysis 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 meta-analysis. These include the study of uncommon adverse outcomes of therapies free of the confounding and bias of non-experimental studies, 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, the combination of studies involved in the approval process for new therapies, and the study of positive effects of therapies, as in the investigation of new indications for existing therapies, particularly when the outcomes being studied are uncommon or the past studies have been small.

One major challenge to the investigation of adverse events using non-experimental studies involves obtaining information on these events that is unconfounded by indication (see Chapter 37). These adverse events often occur only rarely, making their evaluation still more difficult. The results of non-experimental 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. Meta-analysis, by combining results from many *randomized* studies, can address the problem of rare events and rectify the associated lack of adequate statistical power in a setting free of the confounding and bias of non-experimental studies. When reports of several investigations of a specific suspected adverse drug reaction disagree, whether randomized or non-experimental in design, meta-analysis can also be used to help resolve 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, other differences in protocols, or a host of other reasons possibly related to the susceptibility of the 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.

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.

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.[33]

Among the key recommendations from SPERT was that sponsors plan a series of repeated, cumulative meta-analyses of the safety data obtained from the studies conducted within the development program. Leading up to these meta-analyses, sponsors need to develop clear definitions of adverse events of special interest and to standardize various aspects of data collection and study design, to facilitate combining studies and the interpretation of the combined analyses.

By following a proactive approach during development, including periodic updating of cumulative meta-analyses, 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. 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. Meta-analysis can be used to explore these questions with improved statistical power.

The use of meta-analysis 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, meta-analysis could have been used to summarize evidence prior to embarking on a very large-scale, multicenter, randomized trial.[34]

Evidence-based medicine requires the use of the best evidence available in making decisions about the care of patients. Traditional meta-analyses, 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 health-care 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 pharmacological treatment compares with other available pharmacological treatments in terms of safety and efficacy for the specific condition (see Chapter 32).

Extended meta-analytic techniques such as indirect comparisons[35] and multiple treatment meta-analyses can combine all available evidence in a single analysis.[36] 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 below, are met.[37]

Methodologic problems to be addressed by pharmacoepidemiologic research

As the skeptical reader might imagine, many methodologic issues can arise in the context of performing a meta-analysis. 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 meta-analysis used the term "study quality." More recent efforts (e.g., PRISMA[38]) have adopted language that refers to susceptibility to bias or likelihood of bias. We adopt that new terminology in the remainder of this chapter. Meta-analysis 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.[5] 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 meta-analytic process to a different, and potentially serious, form of bias.

Combinability of studies

Clearly, no one would suggest combining studies that are so diverse 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. 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 non-randomized studies be combined with randomized studies? Should non-randomized studies ever be used in a meta-analysis? 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.

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.[39–49] The usual pattern is that statistically significant results are more likely to be published than non-significant results, although this bias may not be as severe for randomized studies as it is for non-randomized studies.[41,50,51] While one could simply decide not to include unpublished studies in a meta-analysis, since those data have often not been peer-reviewed,[52] unpublished data can represent a large proportion of all available data.[53] This may increasingly be the case, given the availability of results on the website clinicaltrials.gov. 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 meta-analysis 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).

Publication bias is a potentially serious limitation to any meta-analysis. For example, Sutton and colleagues[54] found that in four of 48 meta-analyses they examined, there was evidence that the statistical inferences would have changed after the overall effect estimate was adjusted for publication

bias. The retrospective identification of completed unpublished trials is clearly possible[53] in some instances, but generally is not practical. One study[55] used a survey of investigators to attempt to identify unpublished studies. The authors surveyed 42 000 obstetricians and pediatricians, asking whether they had participated in any unpublished trials completed more than 2 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.[42] These include reference bias, that is preferential citation of significant findings;[56] language bias, that is exclusion of studies in languages other than English;[57,58] and bias related to source of funding.[59–61] These related biases have been termed "dissemination bias" by Sutton and colleagues, who found that the threat of such biases is more severe in non-randomized studies of an intervention.[51]

Considerable efforts to reduce publication bias have been made. To be considered for publication, many journals require that clinical trials were publicly registered prior to participant enrolment.[62] In addition, the FDA Amendments Act of 2007 (FDAAA PL 110-85) requires that protocols for all clinical trials involving FDA-regulated drugs and biologics be registered in a publicly accessible registry. This law also requires the registration of trial results within 1 year of trial completion.[62] All these efforts will likely lead to substantial reduction of the impact of publication bias. However, there are concerns about the quality of the results and the possibility that study findings could be misrepresented or misinterpreted, since registration of trials results is subject to less scrutiny than publications in peer-reviewed journals.

There is room to decrease publication bias even further: calls to simplify access to FDA and other regulatory agency reviews, and to create links from such reviews to literature search engines such as MEDLINE have been made.[62]

Bias in the abstraction of data

Meta-analysis, 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.[63] In a meta-analysis of gastrointestinal side effects of NSAIDs,[64] Chalmers and colleagues examined over 500 randomized studies. 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.[63]

In a number of instances, more than one meta-analysis has been performed in the same general area of disease and treatment. A review of 20 of these instances[52] showed that, for almost all disease/treatment areas, there were differences between two meta-analyses of the same topic in the acceptance and rejection of papers to be included. 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 meta-analyses 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 meta-analysis an objective, reproducible activity, there is evidently some judgment involved.

In a separate commentary, DerSimonian[65] reanalyzed data from one meta-analysis and one clinical review of parenteral nutrition with branched-chain amino acids in hepatic encephalopathy. She pointed to differences in the data extracted by the two sets of authors[66,67] 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.

Currently available solutions

This section will first present the general principles of meta-analysis and a general framework for the methods typically employed in a meta-analysis. Since much of the general framework for conducting systematic reviews and explanation of the methods typically employed in a meta-analysis have been presented in review articles in major clinical journals,[8,9,68] 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 meta-analyses. This statement is an update of a previous guideline, known as QUOROM. It consists of a 27-item checklist and a flow diagram. It describes the rationale for including each one of the items with supporting references and provides examples of good reporting.[38] The flow diagram describes the number of studies at each phase of the meta-analysis, 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 meta-analysis of observational studies is available as well.[25]

Another very useful source of information on how to conduct systematic reviews and meta-analysis is the Cochrane Handbook for Systematic Reviews.[69] 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.

Table 40.1 General steps involved in conducting a meta-analysis

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

Steps involved in performing a meta-analysis (Table 40.1)

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 meta-analysis. A well-formulated question should have a clearly defined patient problem, intervention, comparator, and outcome of interest. This framework is called PICO and stands for patient, intervention, comparison, and outcome.[70] The important primary question might be "Are NSAIDs used for the treatment of pain associated with an increased risk of gastrointestinal side effects, 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. 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

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.[71–73] Use of search terms

that are too non-specific 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 non-relevant studies).[74] Nonetheless, one term "random*[tw]" is able to retrieve all randomized controlled trials (RCTs) and improves the specificity of the search strategy to 29%.[74] 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 non-relevant 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, meta-analyses, reviews, or practice guidelines. It has been shown that this approach reduces by 20% the number of non-relevant citations, without losing any of the relevant trials.[75]

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

A set of rules for including and excluding studies from the meta-analysis should be defined during the planning stage of the meta-analysis 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 meta-analysis 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 non-randomized studies.

In establishing inclusion/ exclusion criteria, one is also necessarily defining the question being addressed by the meta-analysis. 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 the examination of the effects of research design on outcome (e.g., do randomized and non-randomized studies tend to show different effects of therapy?) or the exploration of subgroup effects. As an example, in a meta-analysis of aspirin administered following myocardial infarction, restriction of the meta-analysis 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.[10] 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 meta-analyses 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 meta-analysis is motivated, in part, by knowledge of the results of the

component studies, is a more subtle, and perhaps more important, question.)

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 meta-analysis. 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.[76]

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 meta-analysis of the design features and patient characteristics that will be of clinical or academic interest may help avoid over- or under-collecting information. It is generally advisable, when possible, to collect raw data on outcome measures, for example 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 meta-analysis or may have been calculated incorrectly by the original authors.

Many articles on "how to do a meta-analysis" (e.g., Sacks *et al.*[8] L'Abbe *et al.*[9]), and the PRISMA guidelines, recommend that the meta-analyst assess the quality of the studies being considered in a meta-analysis. 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 meta-analyses 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.[77,78] Several examples of quality evaluation systems that have been proposed may be of interest.[79,80] Issues related to quality scoring have been discussed more generally by Moher and colleagues,[81] and an annotated checklist of quality scoring systems is available.[82]

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.[83,84] Jüni and colleagues,[84] for example, examined studies comparing low molecular weight heparin with standard heparin with respect to prevention of postoperative thrombosis. 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 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, for example whether or not the assessment of outcome is blinded to treatment status.

Thus, in a given meta-analysis, one might wish to examine *specific* aspects of study design that are unique to that clinical or statistical situation.[83–86] For example, Schulz and colleagues[86] 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. This specific finding was not detected when these same authors looked for an overall association between quality score and treatment effect. In the analysis of low molecular weight heparin, Jüni and colleagues[84] 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. 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[87] suggest the use of statistical models to investigate the association between specific design factors and study findings. 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 of 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 that authors of systematic reviews assess six domains to determine the quality of each one of the studies included in the analysis.[88] These domains are: 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 meta-analyses in which modest but important inter-reader variability has been demonstrated.[63,64] 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.[52] 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 issue of the effect of masking on the results of meta-analyses.[89] This study compared the results of the same meta-analyses performed independently by separate teams of meta-analysts, with one team masked and the other unmasked. The masked and unmasked teams produced nearly identical results on a series of five meta-analyses, lending little support to the need for masking.

Perform statistical analyses

OR, RR, or RD, does it matter?

There are three summary measures of effect size that can be used in meta-analysis when the outcome of interest is binary (e.g., proportion of subjects with pain relief): relative risk (RR), odds ratio (OR), or risk difference (RD). Although, the summary measure used does not affect the statistical significance of the results,[90] the choice of effect measure could affect the transferability of results of the meta-analysis 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.[91]

RR and RD are easier to interpret than ORs. In general, probabilities are more intuitive than odds. When the baseline (untreated) risk is constant across studies, or assumed fixed, 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 meta-analysis is that, in an empirical study of a large number of meta-analyses, RDs displayed more heterogeneity than ORs, that is the results from study to study appeared more inconsistent with RDs. 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.[91] RR and OR are more consistent than RD[90,91] and therefore are preferred from this perspective. There was no difference in heterogeneity, in this same sample of meta-analyses, on average between RR and OR.[91]

ORs have better mathematical properties than RRs. For example switching the roles of the event and non-event in the analysis is of no consequence for ORs; the new OR is the reciprocal of the original odds ratio (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 meta-analysis the effect of this reversal cannot be predicted.[91]

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.*[92]

Choice of statistical test

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 estimates of relative risk across studies, for example, some form of weighted average of within-study results is appropriate, and several of these exist.[93] 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.[28,94] Other approaches include inverse-variance weighted averages of study-specific estimates of multivariate-adjusted relative risks and exact stratified odds ratios.[93]

Bias in statistical methods is also discussed by Tang,[95] who shows that inverse-variance methods may introduce bias in meta-analyses of binary outcomes. 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 meta-analysis, 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.[96]

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 of 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, for example 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 effect.[97,98] 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.[99–102]

The practical consequence of the random-effects models is to produce wider confidence intervals than would otherwise be produced by the traditional methods.[97,98] 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.[85]

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.[97] 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- versus random-effects models is presented by Hedges and Vevea.[103] Villar and colleagues[104] compared results of fixed- and random-effects models on an empirical basis. 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-effects models applied to the same data. Explanations for this phenomenon are considered in the section on publication bias, below.

Bayesian statistical methods are also being proposed with increasing frequency in the statistical literature.[105–108] These methods can incorporate into the analysis the investigator's prior beliefs about the size of an effect or about the factors biasing the observed effects. When the investigator has no prior beliefs about the effect, the results of the observed studies are sometimes used to estimate the components of the "prior" distribution. Thus, the final answers reflect the observed data very closely. In practice, when the investigator does not specify prior beliefs, the summary results are similar to those from standard methods, especially the random effects models described above, except that, generally speaking, confidence intervals produced by Bayesian methods will be still wider than those produced by other methods.

Combinability of results from diverse studies: heterogeneity

The underlying question in any meta-analysis 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.[109] 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[110] nicely summarize issues related to potential sources of heterogeneity. 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.[3,111–113] These might be important when various components of study design are correlated with each other, acting as potential confounders. Graphical methods for meta-analysis have also been proposed, that focus on issues related to heterogeneity.[114,115]

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.[116,117] 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 data set 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.[118–120] 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 meta-analytic 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 meta-analyses;
- it does not inherently depend on the number of studies in the meta-analysis; and
- 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).

In recent years, increasingly sophisticated (and complex) approaches to the statistical modeling of heterogeneity have been proposed. Thompson and Sharp,[121] for example, compared different forms of weighted normal errors regression and random effects logistic regression. Hardy and Thompson reviewed regression methods to investigate heterogeneity.[111]

A topic of recurring interest in the meta-analysis literature has been the investigation of risk in the control group as a predictor of treatment benefit. The question generally being addressed is whether high-risk patients benefit more from therapy than low-risk patients. However, the question is analyzed using the estimated risk in the control group as an indicator of risk in the *population* under study. However, risk estimates for individual patients are generally not available in such analyses.[122–124] One exception to the lack of patient-level risk estimates is a paper by Trikalinos and Ioannidis,[125] who present methods for modeling study findings as a function of the risk in the control group, using individual patient data. Sharp and Thompson[122] present a Bayesian approach to this problem, based only on group-level data. They recommend a method relying on the underlying binomial distribution of the binary outcomes. That is, this method is analogous to a logistic regression, in which the *risk* is modeled as a function of treatment status and a study-level summary of patient characteristics.

This differs from other approaches based on modeling the treatment effects.[126]

It has been argued that because of the potential for bias in observational epidemiologic studies, exploring heterogeneity should be the main point of meta-analyses of such studies, rather than producing a single summary measure.[6,85,127]

As an example of the type of analysis that could be used to investigate study design issues, Hennessy and colleagues[128] performed a meta-analysis of non-experimental 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. 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 meta-analysis, 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 meta-analytic 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 meta-analysis can address the problems of rare events. However, the analysis of rare events in meta-analysis is still challenging. Many of the methods to combine data in meta-analysis 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.[129,130] The results of these studies show that fixed effect models should be used over random effect methods[130] 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 meta-analyst 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.[129]

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 two-by-two 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.[130] Smaller continuity corrections such as the reciprocal of the sample size of the opposite treatment arm, in contrast with the traditional continuity correction, produce unbiased results.[130]

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 produces unbiased 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).[129,130] The Bayesian methods often use prior distributions that are non-informative, so as not to impose an assumption about the anticipated effect, and use Markov chain Monte Carlo techniques that are capable of including trials that have no events in one of the arms. However, this method excludes studies with no events in any arm as well. The Bayesian fixed effect models produce unbiased results independently of the imbalance in the allocation of treatment groups[130] and therefore are recommended when dealing with rare events.

When a meta-analysis of rare events is contemplated, a thorough sensitivity analysis using different methods to combine studies is recommended and the results of such analyses should be reported so that the readers can assess the robustness of the results.[131]

Other considerations

A number of somewhat specialized statistical issues have been addressed in recent years. These include how to include both parallel and crossover trials in a single meta-analysis,[132–135] 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 meta-analyses,[136] converting odds ratios to effect sizes so that studies with dichotomous outcomes may be combined directly with studies having continuous outcome measures,[137] and the analysis of single patient (*N*-of-1) trials to estimate population treatment effects and to evaluate individual responses to treatment.[138] Nam and colleagues[139] discuss the analysis of studies with multiple, correlated outcomes. Recently published work of particular interest to epidemiologists includes the analysis of dose–response data from epidemiologic data,[140,141] a method for combining disparate designs (case–control, comparative cohort, and uncontrolled cohort studies),[142] and exact methods for case–control and follow-up studies.[143]

Formulate conclusions and recommendations

As with all research, the conclusions of a meta-analysis should be clearly summarized, with appropriate interpretation of the strengths and weaknesses of the meta-analysis. 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 meta-analysis should be stated as such, and not as conclusions.

Approaches to selected methodologic problems in meta-analysis

Publication bias

As discussed above, when the primary source of data for a meta-analysis 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 non-significant findings. A practical technique for determining the potential for publication bias is the "funnel plot," first proposed by Light and Pillemer.[144] 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 meta-analyst should also evaluate the ethics of presenting a summary result that is likely to represent an overestimate of the effect in question.

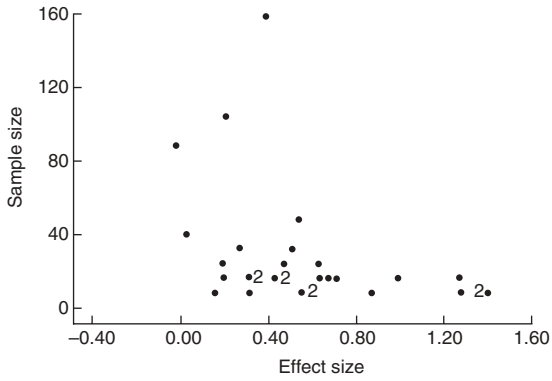


Figure 40.1 Funnel plot for published studies only: analysis of data from Devine and Cook's review of psychoeducational programs for surgical patients.[145]
Reprinted by permission of the publishers from Summing Up: The Science of Reviewing Research, by Richard J. Light and David B. Pillemer, Cambridge MA: Harvard University Press.[144] Copyright 1984 by the President and Fellows of Harvard College.

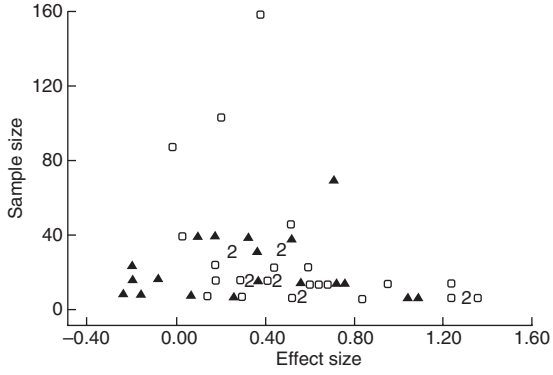


Figure 40.2 Funnel plot for published studies (open boxes) and unpublished (closed triangles) combined only: analysis of data from Devine and Cook's review of psychoeducational programs for surgical patients.[145]
Reprinted by permission of the publishers from Summing Up: The Science of Reviewing Research, by Richard J. Light and David B. Pillemer, Cambridge MA: Harvard University Press.[144] Copyright 1984 by the President and Fellows of Harvard College.

Two examples of funnel plots are given in Figures 40.1 and 40.2. These plots represent studies of psychoeducational programs for surgical patients.[144,145] 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[146] 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. 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.[147] A similar point is made by Terrin and colleagues.[148]

Several mathematical approaches to the problem of publication bias have been proposed. An early method, first described by Rosenthal,[149] is the calculation of a "fail-safe *N*" when the result of the meta-analysis 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 meta-analysis 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 non-significant. 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,[150,151] formal methods to test for the presence of publication bias[152–154] and methods to adjust summary estimates to account for unpublished studies,[150,155–157] 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,[158] 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 above, 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 data set, 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-effects 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[104] found a similar phenomenon in studies in perinatal medicine.

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.[159] Others have suggested obtaining unpublished data from the Food and Drug Administration (FDA), an approach used by Turner *et al.*[160] These authors obtained reviews from 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.

Going one step further, prospective meta-analyses can be conducted.[161,162] These are meta-analyses 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 meta-analysis is presented below.

Indirect comparison and simultaneous comparison of treatments available for specific conditions

What heath-care 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 (see Chapter 32), 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.

Indirect evidence involves using data from trials that have compared medication "A" versus medication "B", and from trials that have compared medication "A" versus medication "C", to draw conclusions about the effect of medication "B" relative to medication "C" (Figure 40.3). It is crucial that when an indirect comparison is estimated, the

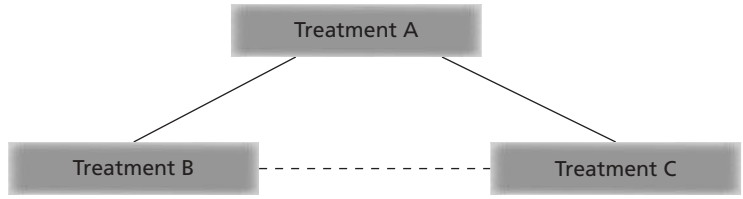


Figure 40.3 Indirect evidence involves using data from trials that have compared medication "A" versus medication "B", and from trials that have compared medication "A" versus medication "C", to draw conclusions about the effect of medication "B" relative to medication "C" (dotted line).

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.[37] 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 meta-analytic 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).

There is, however, a cost in terms of precision. Specifically, indirect evidence estimates are less precise than direct estimates because the variance of the indirect comparison of B versus C is the sum of the variances of the two comparisons estimated above (A vs. B and A vs. C).[163]

When what is needed is a comparison of all available treatments for a specific condition, more flexible analyses are appropriate. An extended meta-analytic method, such as a mixed treatment comparison, a network meta-analysis, or a multiple treatment meta-analysis permit the pharmacoepidemiologist to perform simultaneous comparisons of all treatments. The treatments compared should have a common comparator or need to be otherwise "connected" (Figures 40.4 and 40.5). These techniques often use a Bayesian framework, not because Bayesian input (prior information) is needed, (in fact the priors are generally specified as non-informative), but because of the flexibility of the software.

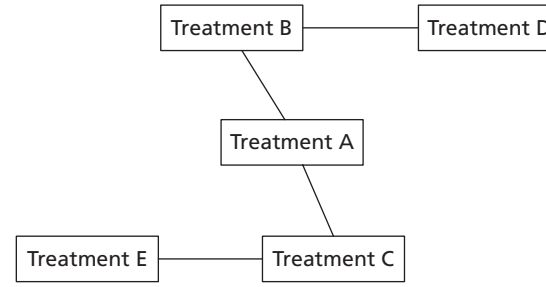


Figure 40.4 Example of a network of treatments. In this case all treatments are connected.

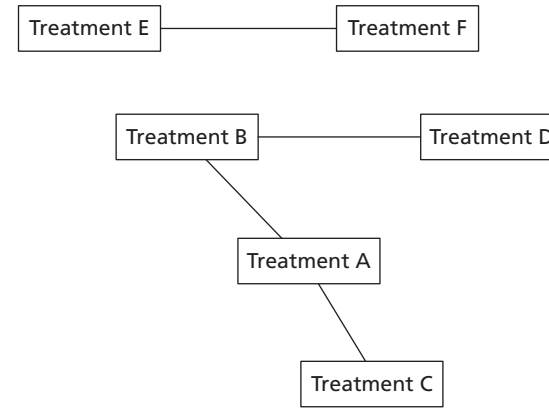


Figure 40.5 Example of a network of treatments. In this case treatments are not connected. Treatments E and F are not connected with the other treatments and therefore they cannot be part of the indirect comparisons.

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.[164] 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 meta-analysis.

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 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.[37]

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 meta-analytic techniques can adjust for possible imbalances of such effect modifiers by incorporating these variables into the statistical model.[165,166] 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.

Case studies of applications of meta-analysis

Investigation of adverse effects

As mentioned earlier, the investigation of adverse or unwanted effects of existing therapies is an important application of meta-analysis. 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. Meta-analysis 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 meta-analysis.[167]

The assessment of the cardiovascular safety of rosiglitazone, a medication used to lower blood glucose, provides an excellent example of a situation in which meta-analysis has been helpful. 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 meta-analysis was conducted.[168]

The authors of this meta-analysis searched published literature, the FDA web site, 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 (Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of glycemia in Diabetes). The RECORD study was designed to evaluate the cardiovascular safety of rosiglitazone. FDA presented the results of this review at an advisory committee meeting in July of 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.[169]

This meta-analysis illustrates some of the challenges researchers face when performing meta-analysis: 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,[129] 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.[170] 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 to report any discrepancies.

New indications for existing therapies

Meta-analysis 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.[171]

To determine the efficacy of antidepressants in the treatment of fibromyalgia a meta-analysis of randomized controlled clinical trials was conducted by Hauser and colleagues.[172] The authors searched MEDLINE, PsycINFO, Scopus, the Cochrane Library databases, and reference sections of original studies, meta-analyses, 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 8 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 the effect sizes 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 meta-analysis illustrated the utility of meta-analysis 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 meta-analysis 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 meta-analysis.[173] 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 (non-fatal). 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 non-behavioral 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. 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 non-psychiatric 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 to 64, and reduced in those aged 65 and older. For suicidal behavior or ideation, the ORs were 1.62 (95% CI: 0.97–2.71) for participants aged less than 25, 0.79 (95% CI: 0.64–0.98) for those aged 25–64, and 0.37 (95% CI: 0.18–0.76) for those aged 65 and older. For suicidal behavior only, for the same age groupings, the ORs were 2.30 (95% CI: 1.04–5.09), 0.87 (95% CI: 0.58– to 1.29), and 0.06 (95% CI: 0.01–0.58), 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 less than 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.[173]

This meta-analysis nicely illustrates the amount of effort and regulatory authority 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 meta-analysis shows the power of individual data meta-analysis to identify subgroups of patients at higher risk of developing adverse events, and the process of adjudicating adverse events that need to be followed when the outcome of interest has not been prespecified in the trials or has not been reported in publications.

Event adjudication is thought to increase the reliability of assessing the treatment effects on outcomes by discarding events that are not valid and, therefore, it is used to reduce bias and increase precision.[174] However, the process of adjudicating events is resource intensive, cumbersome, and costly.

In a different clinical context, however, a study that assessed the benefits of event adjudication showed no clear benefit.[174] The events assessed were myocardial infarction, stroke, and cardiovascular death. The ORs after adjudicating events were similar to the ORs before adjudicating events. This finding of no clear benefit could not be generalized easily to events that are harder to diagnose, such as subtype of strokes instead of overall strokes or, as in the antidepressant meta-analysis, suicidality.[174]

Saving time and resources if you believe a meta-analysis

One of the potential benefits of meta-analysis is the ability to shorten the time between a medical research finding and the clinical implementation of a new therapy. 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 meta-analysis in the approval context, Webber and colleagues[175] reported the use of meta-analysis of ECG data from several clinical pharmacology studies for two drug application submissions. They calculated a pooled estimate for the difference between active doses and placebo in 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 meta-analysis," that is performing a new meta-analysis each time the results of a new clinical trial are published.[34,176] Antman *et al.*[34] applied this

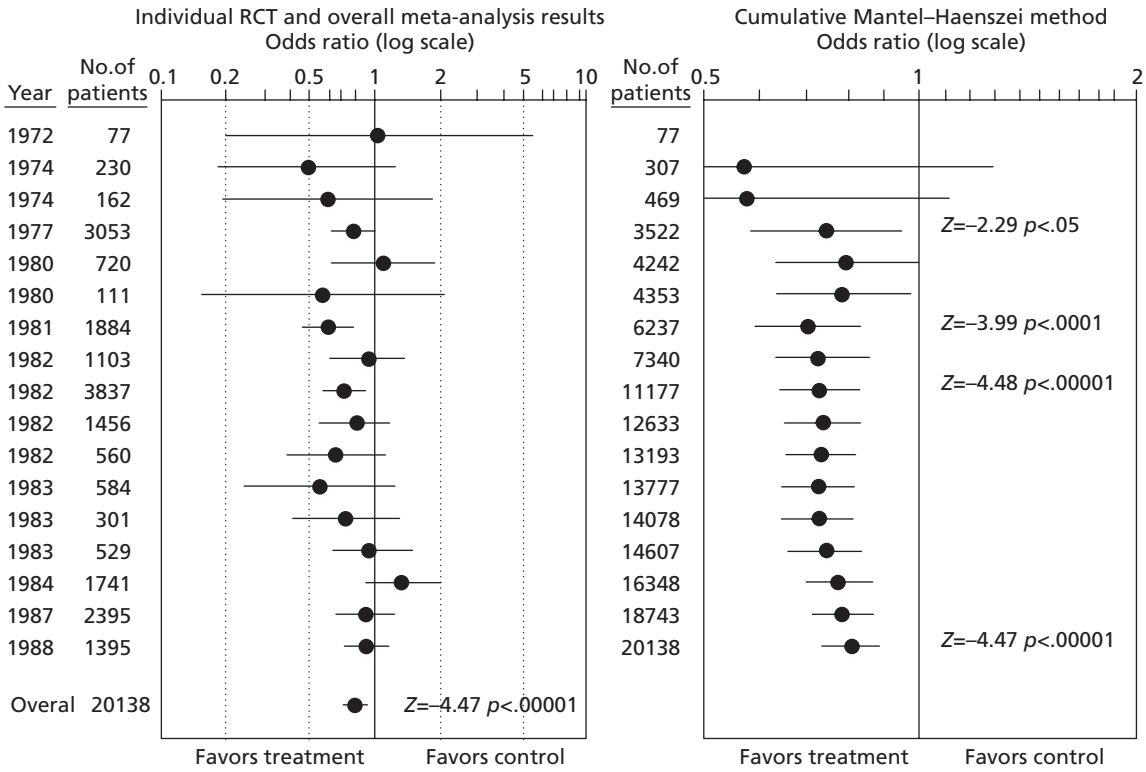


Figure 40.6 Results of 17 randomized control trials of the effect of oral beta-blockers for secondary prevention of mortality in patients surviving a myocardial infarction presented as two types of meta-analyses. On the left is the traditional one, revealing many trials with non-significant results but a highly significant estimate of the pooled results on the bottom of the panel. On the right, the same data are presented as cumulative meta-analyses, 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.[34] Reproduced from Antman *et al.*[34] with permission from the American Medical Association.

technique in combination with a classification scheme of the treatment recommendations for myocardial infarction found in review articles and textbook chapters. 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.[34] In the left-hand side of Figure 40.6, reproduced from their paper, the data are presented as a traditional meta-analysis, 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 40.6, the same data are presented as a cumulative meta-analysis, with an updated summary estimate calculated after the completion of each new trial. The cumulative meta-analysis 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 meta-analyses. The issue of multiple statistical tests, for example, is considered by some to be an important consideration. The problem is that testing and estimation procedures may need to make adjustments for the increased probability of a spurious positive finding (type I, or α, error)

introduced by the use of repeated statistical tests.[177] At the least, one might wish to consider using a more stringent criterion for statistical significance than the traditional $p < 0.05$ cutoff. A recent paper proposes a correction to p-values in the context of cumulative meta-analysis.[178]

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 re-evaluate therapies as other treatment strategies evolve for the same conditions.

A final caution with regard to interpreting cumulative meta-analyses relates to the continuing need for well-designed randomized controlled trials. New indications for existing therapies, for example, are often suggested by non-experimental studies, including cohort and case–control studies and non-randomized 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.*[179] 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.[180] For the reasons just outlined, the role of cumulative meta-analysis to demonstrate effectiveness of a therapy in a new indication has not been clarified in actual regulatory settings. Specifically, whether a meta-analysis 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 meta-analyses, prior to knowing the results of the component studies, may be useful.

Cumulative meta-analysis as a tool to detect harm signals earlier

Cumulative meta-analysis could be used as a tool to detect safety signals earlier. Rofecoxib, a cyclo-oxygenase-2 inhibitor, was withdrawn from the market in September, 2004 because of cardiovascular adverse effects. A cumulative meta-analysis 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.[181]

The authors identified 18 randomized controlled trials and found that by the end of 2000 (4 years before the withdrawal), the relative risk was 2.30 (95% CI: 1.22–4.33), and 1 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, non-naproxen 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 meta-analysis for the evaluation of safety signals brings to light potential methodologic problems that are shared by traditional meta-analysis. 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 non-naproxen 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.[182]

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. A more extensive discussion of the multiplicity issue in safety assessments is presented by Crowe and colleagues in the context of drug development.[33] Additional references can be found in that paper, as well.

Fourth, it is uncertain whether cumulative meta-analysis 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 meta-analysis would not always be expected to detect harms earlier.

Indirect comparisons 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 meta-analyses. Authors of this meta-analysis 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 8 weeks, or between 6 and 12 weeks when data at 8 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 8 weeks of treatment.

The authors calculated the odd ratios (OR) 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 (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 health-care 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.[183] 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.[184] Therefore, a meta-analysis that relies primarily on published data, as this study did, will overestimate the effect size of treatments.

FDA's regulatory role

In recent years, the FDA has used meta-analysis to investigate adverse events associated with the use of certain drugs. The findings from those meta-analyses 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.[185] 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 meta-analysis 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 for continued use in patients. A decision may be made either to withdraw the drug completely, or to withdraw its use for a particular indication. For example, meta-analysis 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 meta-analysis published by Yahav *et al.*, based on 38 clinical trials.[186] The FDA performed its own meta-analysis 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 meta-analysis 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 meta-analysis, 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 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 meta-analysis of rosiglitazone on a patient-level data as well (Advisory Committee, June, 2007). For this meta-analysis, the database for the FDA reanalysis differed on 14 studies as 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 meta-analysis showed that the overall odds ratio 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 increase in risk in the studies comparing rosiglitazone with metformin or a sulfonylurea.

The FDA continues the effort to review both rosiglitazone and cefepime by using prospective meta-analysis. By promoting early communication (i.e., prior to knowing the outcome of the studies) to adjudicate outcomes or events of interest in a blinded fashion that is defined in common among sponsors, prospective meta-analysis may also help reduce publication bias in the long run. Further, the 2007 FDA Amendment Act (FDAAA PL 110-85) mandates that investigators and sponsors register all clinical trials, with the exception of Phase I trials, on the government-sponsored website clinicaltrials.gov. In the FDAAA PL 110-85, a clinical trial is defined as "any research study that

prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes." Such continuous effort will help further reduce the potential publication bias in conducting meta-analysis. Notably, this mandate covers both industry-funded studies and all others.

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[33] outlined broad principles, and pointed toward potential solutions, including the use of a tiered approach to defining adverse events. 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 meta-analyses 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.

When hundreds of categories of events are tabulated, it's 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 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.[187,188]

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 meta-analytic evidence as helping to establish effectiveness. It is less obvious what to do with the results of a meta-analysis when there is substantial heterogeneity. If the heterogeneity is adequately explained in the analysis in terms of subgroup effects, or trial quality, meta-analysis 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 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 32), 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 compari-

sons, 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.[189] How and when to incorporate studies that are not head-to-head comparisons needs further empirical study.

The inclusion of non-experimental observational studies in meta-analyses, 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, FDA has been charged with the goal of establishing a network of observational databases, to be known as the Sentinel Network, aimed at exactly this type of assessment of drug safety in clinical practice (described in detail in Chapter 30). The legislative mandate is to provide access to claims or electronic medical records data from 100 million individuals by 2012.[190,191] Current information regarding the Sentinel Initiative, which will evolve into the Network, can be accessed at the Sentinel website.[192] Clearly, issues of bias and confounding will need to be addressed as the Sentinel Initiative moves forward. A pilot version of the Sentinel Network, known as "Mini-Sentinel," is currently underway.[193]

In a potentially related effort, a public–private partnership, known as the Observational Medical Outcomes Partnership (OMOP), funded by the pharmaceutical industry, and including representatives from industry, FDA, and academic institutions, has been conducting methodologic research to determine which approaches to analysis of such observational data provide the "best" (least biased, most consistent, most precise) results (see Chapter 30). The current OMOP approach, and that likely to be used by the Sentinel Network, is that of a distributed network. That is, providers of data will house the data, and will provide analytical results at the aggregate-level only, to a central group that will evaluate 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. Providers of data to the OMOP core group have all agreed, as a condition of participation, to adopt a common data model, and such a model has been implemented. A series of presentations and several articles (under review at the time of this writing) are available on the OMOP website.[194]

In conclusion, while there are no easy answers to many of the questions presented above, it is clear that meta-analysis will play an increasingly important role in the formulation of treatment and policy recommendations. Thus, the quality of the meta-analyses 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. Meta-analyses, if they are carefully interpreted in view of their strengths and weaknesses, should prove to be extremely helpful in pharmacoepidemiologic research.

References

1. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986; **7**: 177–88.
2. Last JM. *A Dictionary of Epidemiology*, 2nd edn. New York, NY: Oxford University Press, 1988.
3. Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. *Epidemiology* 1993; **4**: 218–28.
4. Feinstein AR. Para-analysis, faute de mieux, and the perils of riding on a data barge. *J Clin Epidemiol* 1989; **42**: 929–35.

5. Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *J Clin Epidemiol* 1997; **50**: 223–9.
6. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; **316**: 140–4.
7. Buffler PA. The evaluation of negative epidemiologic studies: the importance of all available evidence in risk characterization. *Regul Toxicol Pharmacol* 1989; **9**: 34–43.
8. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; **316**: 450–5.
9. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987; **107**: 224–33.
10. Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. *Am J Epidemiol* 1995; **142**: 371–82.
11. Berlin JA. Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *Am J Epidemiol* 1995; **142**: 383–7.
12. Bero LA, Rennie D. The Cochrane Collaboration—preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995; **274**: 1935–8.
13. Doyle J, Waters E, Noonan E, Royle N, Kowalski K, Gill C, *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 (Oxf)* 2010; **32**: 138–40.
14. Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. *Epidemiol Rev* 1992; **14**: 154–76.
15. Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987; **106**: 485–8.
16. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988; **138**: 697–703.
17. Chalmers I. Improving the quality and dissemination of clinical research. In: Lock S, ed. *The Future of Medical Journals: in Commemoration of 150 Years of the British Medical Journal*. London, England: British Medical Society, 1991, pp. 127–46.
18. Kass EH. Reviewing reviews. In: Warren KS, ed. *Coping with the Biomedical Literature: a Primer for the Scientist and the Clinician*. New York, NY: Praeger, 1981, pp. 71–91.
19. Huth EJ. Needed: review articles with more scientific rigor. *Ann Intern Med* 1987; **106**: 470–1.
20. Canner PL. An overview of six clinical trials of aspirin in coronary heart disease. *Stat Med* 1987; **6**: 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**: 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**: 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**: 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**: 1077–83.
25. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, *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**: 2008–12.
26. Berlin JA, Colditz GA. The role of meta-analysis in the regulatory process for foods, drugs, and devices. *JAMA* 1999; **281**: 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**: 1301–9.
28. Greenland S, Salvan A. Bias in the one-step method for pooling study results. *Stat Med* 1990; **9**: 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**: 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**: 469–75.
32. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, *et al.* Meta-analysis of observational studies in epidemiology—A proposal for reporting. *JAMA* 2000; **283**: 2008–12.
33. Crowe BJ, Xia HA, Berlin JA, Watson DJ, Shi H, Lin SL, *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**: 430–40.

34. 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**: 240–8.
35. Gartlehner G, Moore CG. Direct versus indirect comparisons: a summary of the evidence. *Int J Technol Assess Health Care* 2008; **24**: 170–7.
36. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; **23**: 3105–24.
37. 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.
38. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, *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**: W65–W94.
39. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989; **81**: 107–15.
40. 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** (Suppl): III 236–44.
41. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337**: 867–72.
42. 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.
43. 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**: 374–8.
44. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; **263**: 1385–9.
45. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993; Doc No 50.
46. Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990; **263**: 1405–8.
47. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA* 1990; **263**: 1392–5.
48. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997; **315**: 640–5.
49. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998; **279**: 281–6.
50. 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.
51. 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.
52. 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**: 733–44.
53. Early Breast Cancer Trialists' Collaborative Group. *Treatment of Early Breast Cancer*, Vol. I. *Worldwide Evidence*. New York, NY: Oxford University Press, 1990.
54. Sutton AJ, Duval S, Tweedie R, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000; **320**: 1574–7.
55. 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**: 374–80.
56. Gotzsche PC. Reference bias in reports of drug trials. *BMJ Clin Research Ed* 1987; **295**: 654–6.
57. 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**: 326–9.
58. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, *et al.* Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; **347**: 363–6.
59. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986; **1**: 155–8.
60. Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ* 1980; **280**: 833–6.
61. Cho MK, Bero LA. The quality of drug studies published in symposium proceedings. *Ann Intern Med* 1996; **124**: 485–9.
62. O'Connor AB. The need for improved access to FDA reviews. *JAMA* 2009; **302**: 191–3.

63. Chalmers TC. Problems induced by meta-analyses. *Stat Med* 1991; **10**: 971–9.
64. Chalmers TC, Berrier J, Hewitt P, Berlin J, Reitman D, Nagalingam R, *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**: 9–26.
65. DerSimonian R. Parenteral nutrition with branched-chain amino acids in hepatic encephalopathy: meta analysis. *Hepatology* 1990; **11**: 1083–4.
66. 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**: 1033–42.
67. Erikkson LS, Conn HO. Branched-chain amino acids in hepatic encephalopathy. *Gastroenterology* 1990; **99**: 604–7.
68. Thacker SB. Meta-analysis. A quantitative approach to research integration. *JAMA* 1988; **259**: 1685–9.
69. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.1.0, March 2011. www.cochrane-handbook.org (accessed July 2011).
70. 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.
71. Bernstein F. The retrieval of randomized clinical trials in liver diseases from the medical literature: manual versus MEDLARS searches. *Controlled Clin Trials* 1988; **9**: 23–31.
72. Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. *Controlled Clin Trials* 1985; **6**: 306–17.
73. Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature. A comparison of MEDLARS and manual methods. *Controlled Clin Trials* 1985; **6**: 271–9.
74. 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**: 103–9.
75. 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**: 224–5.
76. Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996; **347**: 1024–6.
77. 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**: 255–65.
78. 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.
79. Mahon WA, Daniel EE. A method for the assessment of reports of drug trials. *Can Med Assoc J* 1964; **90**: 565–9.
80. Verhagen AP, de Vet HCW, de Bie RA, Kessels AGH, Boers M, Bouter LM, *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.
81. 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**: 195–208.
82. 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. *Controlled Clin Trials* 1995; **16**: 62–73.
83. Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994; **140**: 300–1.
84. Jüni P, Witschi A, Block R, Egger M. The hazards of scoring the quality of clinical trials: lessons for meta-analysis. *JAMA* 1999; **282**: 1054–60.
85. Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994; **140**: 290–6.
86. 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**: 408–12.
87. 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.
88. *The Cochrane Collaboration's Tool for Assessing Risk of Bias*, 2010. http://www.ohg.cochrane.org/forms/Risk%20of%20bias%20assessment%20tool.pdf (accessed July 2011).
89. Berlin JA, Miles CG, Cirigliano MD, Conill AM, Goldmann DR, Horowitz DA, *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.
90. 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**: 1707–28.
91. Jonathan JD. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002; **21**: 1575–600.
92. 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.
93. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. New York, NY: Van Nostrand Reinhold, 1982.
94. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990; **1**: 43–6.
95. Tang J-L. Weighting bias in meta-analysis of binary outcomes. *J Clin Epidemiol* 2000; **53**: 1130–6.
96. Chang BH, Waternaux C, Lipsitz S. Meta-analysis of binary data: which within study variance estimate to use? *Stat Med* 2001; **20**: 1947–56.
97. Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989; **8**: 141–51.
98. Fleiss JL. The statistical basis of meta-analysis. *Stat Methods Med Res* 1993; **2**: 121–45.
99. Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. *Biometrics* 1999; **55**: 732–7.
100. Aitkin M. Meta-analysis by random effect modelling in generalized linear models. *Stat Med* 1999; **18**: 2343–51.
101. 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.
102. Senn SJ. The many modes of META. *Drug Inf J* 2000; **34**: 535–49.
103. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Methods* 1998; **3**: 486–504.
104. 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.
105. Carlin JB. Meta-analysis for 2 × 2 tables: a Bayesian approach. *Stat Med* 1992; **11**: 141–58.
106. Larose DT, Dey DK. Grouped random effects models for Bayesian meta-analysis. *Stat Med* 1997; **16**: 1817–29.
107. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995; **14**: 2685–99.
108. Eddy DM, Hasselblad V, Shachter R. An introduction to a Bayesian method for meta-analysis: The confidence profile method. *Med Decis Making* 1990; **10**: 15–23.
109. Simon R. Overviews of randomized clinical trials. *Cancer Treat Rep* 1987; **71**: 3–5.
110. Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002; **21**: 1503–11.
111. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Stat Med* 1998; **17**: 841–56.
112. Berlin JA, Antman EM. Advantages and limitations of metaanalytic regressions of clinical trials data. *Online J Curr Clin Trials* 1994; Doc. No. 134.
113. Vanhonacker WR. Meta-analysis and response surface extrapolation: a least squares approach. *Am Stat* 1996; **50**: 294–9.
114. Walker AM, Martin-Moreno JM, Artalejo FR. Odd man out: a graphical approach to meta-analysis. *Am J Pub Health* 1988; **78**: 961–6.
115. Song F. Exploring heterogeneity in meta-analysis: is the L'Abbe plot useful? *J Clin Epidemiol* 1999; **52**: 725–30.
116. Fleiss JL. Analysis of data from multiclinic trials. *Controlled Clin Trials* 1986; **7**: 267–75.
117. Takkouche B, Cadarso-Suarez C, Spiegelman D. Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *Am J Epidemiol* 1999; **150**: 206–15.
118. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539–58.
119. Higgins J, Thompson SG, Deeks J, Altman DG. Measuring inconsistency in meta-analysis. *BMJ* 2003; **327**: 557–60.
120. 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**: 51–61.
121. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999; **18**: 2693–708.

122. Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Stat Med* 2000; **19**: 3251–74.
123. Schmid C, Lau J, McIntosh M, Cappelleri J. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med* 1998; **17**: 1924–42.
124. Cook RJ, Walter SD. A logistic model for trend in $2 \times 2 \times$ kappa tables with applications to meta-analyses. *Biometrics* 1997; **53**: 352–7.
125. Trikalinos TA, Ioannidis JP. Predictive modeling and heterogeneity of baseline risk in meta-analysis of individual patient data. *J Clin Epidemiol* 2001; **54**: 245–52.
126. Schmid CH. Exploring heterogeneity in randomized trials via meta-analysis. *Drug Inform J* 1999; **33**: 211–24.
127. Davey Smith G, Egger M, Phillips AN. Meta-analysis. Beyond the grand mean? *BMJ* 1997; **315**: 1610–4.
128. 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**: 125–33.
129. 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**: 53–77.
130. 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**: 1351–75.
131. 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**: 116–20.
132. Curtin F, Altman DG, Elbourne D. Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes. *Stat Med* 2002; **21**: 2131–44.
133. Curtin F, Elbourne D, Altman DG. Meta-analysis combining parallel and cross-over clinical trials, II: Binary outcomes. *Stat Med* 2002; **21**: 2145–59.
134. 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.
135. 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.
136. Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. *Stat Med* 2002; **21**: 2971–80.
137. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000; **19**: 3127–31.
138. 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**: 401–10.
139. Nam I-S, Mengersen KL, Garthwaite P. Multivariate meta-analysis. *Stat Med* 2003; **22**: 2309–33.
140. Cook R, Brumback BA, Wigg M, Ryan L. Synthesis of evidence from epidemiological studies with interval-censored exposure due to grouping. *Biometrics* 2001; **57**: 671–80.
141. Shi J, Copas J. Meta-analysis for trend estimation. *Stat Med* 2004; **23**: 3–19.
142. Brumback BA, Holmes LB, Ryan LM. Adverse effects of chorionic villus sampling: a meta-analysis. *Stat Med* 1999; **18**: 2163–75.
143. Martin DO, Austin H. An exact method for meta-analysis of case-control and follow-up studies. *Epidemiology* 2000; **11**: 255–60.
144. Light RJ, Pillemer DB. *Summing Up: the Science of Reviewing Research*. Cambridge, MA: Harvard University Press, 1984.
145. 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.
146. Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *J Clin Epidemiol* 2001; **54**: 1046–55.
147. Sterne JAC, Egger M, Smith GD. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; **323**: 101–5.
148. Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Stat Med* 2003; **22**: 2113–26.
149. Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979; **86**: 638–41.
150. Iyengar S, Greenhouse JB. Selection models and the file-drawer problem. *Stat Sci* 1988; **3**: 109–17.
151. Gleser LJ, Olkin I. Models for estimating the number of unpublished studies. *Stat Med* 1996; **15**: 2493–507.

152. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629–34.
153. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; **50**: 1088–101.
154. Dear KBG, Begg CB. An approach for assessing publication bias prior to performing a meta-analysis. *Stat Sci* 1992; **7**: 237–45.
155. 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.
156. Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* 1995; **60**: 419–35.
157. 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.
158. 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**: 455–63.
159. De AC, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, *et al.* Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med* 2004; **351**: 1250–1.
160. 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**: 252–60.
161. 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**: 1005–8.
162. Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, *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**: 1267–78.
163. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; **50**: 683–91.
164. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010; **29**: 932–44.
165. 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**: 1861–81.
166. 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**: 857–64.
167. Temple RJ. The regulatory evolution of the integrated safety summary. *Drug Inform J* 1991; **25**: 485–92.
168. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007; **356**: 2457–71.
169. FDA. FDA Drug Safety Podcast: FDA significantly restricts access to the diabetes drug Avandia, 2010 September 28. http://www.fda.gov/Drugs/DrugSafety/DrugSafetyPodcasts/ucm227511.htm (accessed July 2011).
170. Bracken MB. Rosiglitazone and cardiovascular risk. *N Engl J Med* 2007; **357**: 937–8.
171. 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.
172. Hauser W, Bernardy K, Uceyler N, Sommer C. Treatment of fibromyalgia syndrome with antidepressants: a meta-analysis. *JAMA* 2009; **301**: 198–209.
173. Stone M, Laughren T, Jones ML, Levenson M, Holland PC, Hughes A, *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.
174. Pogue J, Walter SD, Yusuf S. Evaluating the benefit of event adjudication of cardiovascular outcomes in large simple RCTs. *Clin Trials* 2009; **6**: 239–51.
175. Webber DM, Montague TH, Bird NP. Meta-analysis of QTc interval-pooling data from heterogeneous trials. *Pharm Stat* 2002; **1**: 17–23.
176. 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.
177. Berkey CS, Mosteller F, Lau J, Antman EM. Uncertainty of the time of first significance in random effects cumulative meta-analysis. *Controlled Clin Trials* 1996; **17**: 357–71.
178. Lan KKG, Hu M, Cappelleri JC. Applying the law of iterated logarithm to cumulative meta-analysis of a continuous endpoint. *Statistica Sinica* 2003; **13**: 1135–45.

179. Ziegler RG, Mayne ST, Swanson CA. Nutrition and lung cancer. *Cancer Causes Control* 1996; **7**: 157–77.
180. Hennekens CH, Buring JE, Manson JE, Stampfer M, Rosner B, Cook NR, *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**: 1145–9.
181. Juni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet* 2004; **364**: 2021–9.
182. Gandey A. *Merck Fights Back*. Medscape Medical News, 2005. http://www.medscape.com/viewarticle/538074 (accessed July 2011).
183. Turner E, Moreno SG, Sutton AJ. Ranking antidepressants. *Lancet* 2009; **373**: 1760–2.
184. 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**: 252–60.
185. FDA. *Statistical Review and Evaluation. Antiepileptic Drugs and Suicidality*. http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM192556.pdf (accessed July 2011).
186. 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**: 338–48.
187. Solomon SD, McMurray JJ, Pfeffer MA, Wittes J, Fowler R, Finn P, *et al.* Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. *N Engl J Med* 2005; **352**: 1071–80.
188. 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.
189. 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**: 1–5.
190. 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.
191. US Department of Health and Human Services. The Sentinel Initiative: A National Strategy for Monitoring Medical Product Safety, 2008. http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm089474.htm (accessed July 2011).
192. US Department of Health and Human Services. *FDA's Sentinel Initiative—Ongoing Projects*. http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm203500.htm (accessed July 2011).
193. Platt R. *FDA's Mini-Sentinel Program*, 2010. http://www.brookings.edu/~/media/Files/events/2010/0111_sentinel_workshop/06%20Sentinel%20Initiative%20Platt%20Brookings%2020100111%20v05%20distribution.pdf (accessed July 2011).
194. Foundation for the National Institutes of Health. Observational Medical Outcomes Partnership. http://omopfnihorg/?q=node/22 (accessed July 2011).

CHAPTER 41
Validity of Pharmacoepidemiologic Drug and Diagnosis Data

Suzanne L. West,[1,3] Mary Elizabeth Ritchey,[2] and Charles Poole[3]

[1] RTI International, Research Triangle Park, NC, USA

[2] Food and Drug Administration, Center for Devices and Radiological Health, Silver Spring, MD, USA

[3] Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

Introduction

In discussing the quality of data for research, Gordis remarked that epidemiologists have become so enamored with statistical analysis of the data that they have paid too little attention to the validity of the raw data being analyzed with these sophisticated techniques.[1] Although this statement referred to questionnaire data, it applies equally to data generated by abstracting medical records or data from electronic databases. Whatever the source of the data, the veracity of a study's conclusions rests on the validity of its data.

We begin this chapter by discussing the validity of the drug and diagnosis information used by clinicians in the management of patients' care. Next, we discuss measurement error, describing the different types of error and error detection methods, exploring how errors may affect the point estimate, and describing current techniques for mitigation. In the remainder of the chapter we use two associations to illustrate validity concerns when using data from administrative claims, electronic health records or questionnaire responses: the association between non-steroidal anti-inflammatory drugs (NSAIDs) and gastrointestinal (GI) bleeding, and NSAIDs and myocardial infarction (MI).

Clinical problems to be addressed by pharmacoepidemiologic research

Physicians rely on patient-supplied information on past drug use and illness to assist with the diagnosis of current disease. Proper diagnosis and treatment of current illnesses may be compromised by poor recall of past illnesses and drugs. Patients' recall abilities compromise a physician's ability to diagnose and/or prescribe successfully and may play a role in the success of drug therapy. The patient needs to recall the physician's instructions for most effective drug use. Brody found that 55 (53%) of 104 patients interviewed immediately after seeing their physician made one or more errors in recalling their therapeutic regimens.[2] Patient recall may be even poorer for illnesses and medication use that occurred many years previously.

Of particular concern to the subject of this book is the validity of data on drug exposure and disease

The opinions expressed in this chapter by Mary Elizabeth Ritchey are those of the author and do not necessarily represent the official policies of the US Food and Drug Administration

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

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. As noted, 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 any aids to this recall. Superficially at least, patients cannot be considered reliable sources of diagnosis information either; in some instances they may not even have been told 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, from patients and other sources, but the success of these approaches needs to be considered in detail.

Methodologic problems to be solved by pharmacoepidemiologic research

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 (κ), which is agreement corrected for chance.[3]

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 (a synonym is accuracy). 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 gold standard" has been used.[4]

For a binary exposure or outcome measure, such as "ever" versus "never" use of a particular drug, two measures of validity are used. Sensitivity (also called completeness) measures the degree to which the 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 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 41.1 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 high sensitivity tend to have low specificity, and methods with high specificity tend to have low 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, either sensitivity or specificity may be the more

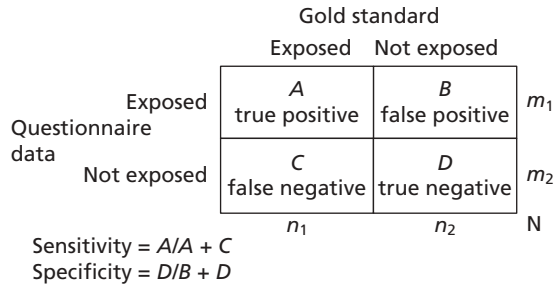


Figure 41.1 Formulas for calculating sensitivity and specificity.

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 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 misclassification.

Because the degree of bias depends on such study-specific conditions as the true prevalence of exposure, no general guidelines can be given. Each study situation must be evaluated on its own merits. For example, suppose in a case–control study that the true odds ratio is OR = 3.0, the sensitivity of an exposure measure is higher among cases (90%) than among controls (80%), the specificity is lower among cases (95%) than among controls (99%), and, for simplification, that the outcome is measured perfectly and there is no control-selection bias. The exposure misclassification will bias the expected effect estimate upward to OR = 3.6 if the true exposure prevalence in the source population is 10%, downward to OR = 2.6 if the true exposure prevalence is 90%, and leave it unbiased at OR = 3.0 if the true exposure prevalence is 70%.[5]

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 and those who are *classified* as having it or not having it. The proportion of persons classified as having the exposure or outcome who are correctly classified is the positive predictive value. The proportion of persons classified as lacking the exposure or outcome who are correctly classified is the negative predictive value. 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 true 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 validity (i.e., 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.

In many validation studies, the confirmation or verification rates are not measures of validity, but merely measures of agreement. In other such investigations, 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 without 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, validation of cases in a case–control study using self-report or administrative data often provides only the positive predictive value that the cases are true cases and does not evaluate the negative predictive value that the controls are truly controls. Ideally, one would design a validation study to calculate sensitivity, specificity, as well as positive and negative predictive values.

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.

For a drug exposure, a true gold standard is a list of all drugs the study participant has taken, 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 drug use as more regular and complete than it actually was, or that use adhered to the typical prescribed regimen. Similarly, substantial gaps may exist between when a prescription is filled and when it is ingested, if it is ingested at all.

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).[6] 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 (Figure 41.2).[7] 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 exposure.

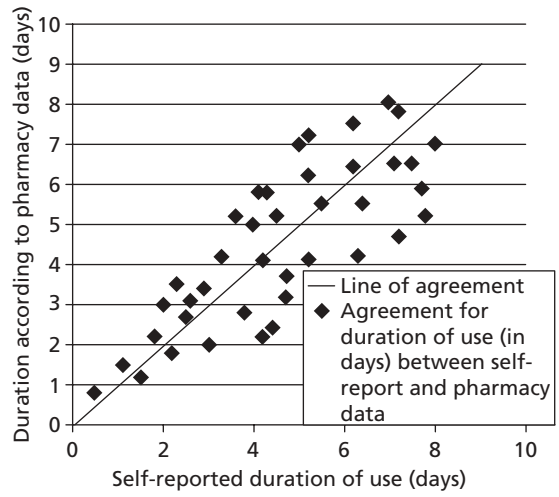


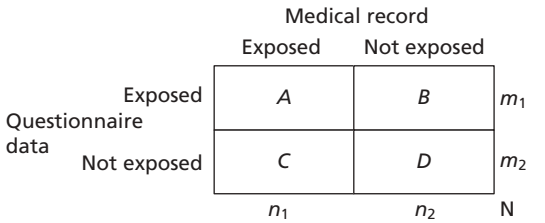
Figure 41.2 Graphic showing line of agreement for continuous variables.

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, though the term is sometimes used, inaccurately, as such.

When different data collection methods or different sources of information are compared (e.g., comparing 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 κ 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.[6,8] The κ statistic is the percentage agreement corrected for chance.[6] Agreement is conventionally considered poor for a



Accuracy = $A + D/N$

Chance agreement (expected) = $((n_1 \times m_1) + (n_2 \times m_2)/N^2$

$$\kappa = \frac{\text{accuracy–chance agreement}}{\text{1–chance agreement}}$$

Figure 41.3 Formulas for calculating the percent agreement and K.

κ statistic less than zero, slight for κ between zero and 0.20, fair for a κ of 0.21–0.40, moderate for a κ of 0.41–0.60, substantial for a κ of 0.61–0.80, and almost perfect for a κ of 0.81–1.00.[3] Figure 41.3 illustrates the percentage agreement and κ 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.[8] It reflects both the average differences in mean values as well as the correlation between measurements. The intraclass correlation coefficient indicates how much of 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 κ.[6]

It is impossible to translate values of measures of agreement, such as κ, into expected degrees of bias in exposure or disease associations.

Measurement error in pharmacoepidemiologic research

Epidemiologic assessments of the effects of a drug on disease incidence depend upon an accurate assessment of both drug exposure and disease occurrence. Measurement error for either factor may 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 using questionnaire data to study the association between drug A and disease B, if some study participants forgot their past exposure to drug A, they would be incorrectly classified as non-exposed. 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.

There are two types of measurement error or misclassification: non-differential and differential.[9] The difference between these errors relates to the variables under study. In particular, differential misclassification occurs when the misclassification of one variable (e.g., drug usage) varies according to the level of another variable (e.g., disease status), so that the direction of the bias is toward or away from the null. For example, in a case–control study of NSAIDs and MI, patients with an MI might recall past NSAID use differently from those who had not had a recent MI. MI cases might ponder the origins of their illness and recall and report NSAID use they otherwise would have forgotten or failed to report (Figure 41.4, cases recall exposure better than controls). Alternatively, patients might be distracted by their illness during the interview and forget their past NSAID use, fail to report it to get the interview over more quickly, or because of psychological denial in favor of something else that they may feel is more likely as an explanation for their disease (Figure 41.4, cases do not recall exposure as well as controls).

Thus, the respondent's state of mind (and possibly that of the interviewer) at the time of the interview determines the overall accuracy of the interview or questionnaire information and the degree to which the accuracy might differ by respondent characteristics (e.g., case or control status). Patients who learn they have serious diseases, and parents who learn the same about their children, often go through phases or stages in

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 41.4 Example of differential misclassification of exposure.

questioning how these illnesses might have come about. In earlier stages, patients often blame themselves. As the time passes, they frequently seek external explanations. The time course of the psychological state of seriously ill patients and their close family members varies highly, but is potentially very important to the validity of interview and questionnaire data they provide. The traditional assumptions that cases remember true exposures better than non-cases (i.e., that exposure classification has higher sensitivity among cases than among non-cases) and that cases intentionally or unintentionally report more false-positive exposures than non-cases (i.e., that exposure classification has lower specificity among cases than among non-cases) are undoubtedly too simplistic for general reliance.

A difference in the accuracy of recall between cases and non-cases could influence the determination of NSAID exposure and the resulting measure of association. In case–control studies, differential misclassification of exposure can result from recall bias.[10] A common belief is that the potential for recall bias can be minimized if the study is designed to obtain complete exposure data, that is information on the names and usage dates for every drug used in the time period of interest.[11]

Non-differential misclassification of exposure occurs when the misclassification of one variable does not vary by the level of another variable and may occur if both cases and controls simply forget their exposures to the same degree. The measure of association is affected by non-differential misclassification of exposure as well; it is usually biased

No Exposure Misclassification

	Exposure				
	Low	Medium	High		
Cases	200	550	700	Medium vs Low	OR = 1.4
Controls	300	600	400	High vs Low	OR = 1.8

0.3 of cases and controls in the high exposure group are misclassified as medium exposure

	Exposure				
	Low	Medium	High		
Cases	200	760	490	Medium vs Low	OR = 1.6
Controls	300	720	280	High vs Low	OR = 2.6

Figure 41.5 Example of non-differential misclassification of exposure when exposure is polychotomous.

toward the null. Exceptions can occur when classification errors are not independent of each other,[12–17] as when participants who are particularly reluctant to report health outcomes that they have experienced are especially unwilling to report medications they have taken as well. Other exceptions to the rule about bias toward the null from non-differential misclassification can occur when there are more than two categories of exposure.[18] We provide a simple hypothetical example using a case–control analysis to illustrate the potential for bias away from the null from independent, non-differential misclassification of an exposure with more than two categories of exposure: low, medium, and high (Figure 41.5). As noted in the figure, if 30% of the cases and controls in the high exposure group are misclassified into the medium exposure group, the odds ratio for medium exposure is relatively unbiased but the odds ratio for high exposure is biased upward to 2.6.

No bias occurs from independent, non-differential misclassification of a binary outcome measure under some circumstances.[19,20] For instance, if no false-positive cases exist, the expected risk ratio will be the risk ratio given correct disease classification multiplied by the ratio of the sensitivity in the exposed group to the sensitivity in the unexposed group. If the sensitivity is independent and non-differential, this ratio equals unity and the risk ratio is unbiased.

Effects of measurement error on the point estimate of association

Copeland *et al.* evaluated misclassification in epidemiologic studies using a series of computer-generated graphs. They showed that the bias—that is discrepancy between the point estimate and the true value of the measure of association—was a function of the disease frequency, exposure frequency, sensitivity, and specificity of the classification.[21] It is instructive to note that Copeland *et al.* were not able to describe bias as a function of the product–moment correlation coefficient, the intraclass correlation coefficient, percentage agreement, or κ. Thus, higher or lower values of these measures, even when one of the measurement methods is a gold standard, should not be interpreted as evidence of greater or lesser degrees of bias. When non-differential misclassification occurred, the point estimate was biased toward the null. Their results for non-differential misclassification of disease also indicated that the rarer the disease, the

more the potential for bias in cohort studies. Likewise, the less prevalent the exposure, the more the potential for bias increases in case–control studies. For differential misclassification, the point estimate could be biased toward or away from the null. This presents a problem for *ad hoc* case–control studies, where recall bias is always a concern.

Copeland *et al.*'s simulations were all done on binary disease and exposure variables. For a continuous variable, non-differential misclassification may not produce a bias towards the null if a perfect correlation exists between the variable as measured and the true value.[8] For example, if both cases and controls in a case–control study underestimate duration of drug use by an equal percentage, a bias towards the null would not occur.

Correcting measures of association for measurement error

Estimates of sensitivity and specificity are required to correct effect estimates for measurement error.[21] These estimates can be derived from previous research or from a subsample within the study analyzed. However, estimates of sensitivity and specificity of exposure classification from previous research are rarely available. Should these estimates be available, they may not be useful since the classification methods need to be similar in both the correctly classified and misclassified data.[22] The classification probabilities will vary according to the questionnaire design, study population, and time period of administration. In addition, the correction methods most familiar to epidemiologists are appropriate for bivariate, not multivariate, data.[23]

For differential misclassification of exposure by disease status (e.g., recall bias), Raphael[24] contends that the researcher is responsible for either presenting a strong case that recall bias did not threaten the study's validity or controlling for it statistically. One approach to estimate the effects of bias is to conduct a sensitivity analysis.[25] 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. 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. In a sensitivity analysis, one alters key assumptions or methods reasonably to see how sensitive the results of a study are to those variations. One key assumption, usually implicit, is that the exposure and the outcome in a study have been measured accurately. With estimates from previous research or "guesstimates" from expert experience and judgment, one can modify this assumption and use a variety of analytic 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.[26,27] Sometimes, wildly implausible degrees of inaccuracy would have to have been present to produce observed associations.

For many years, this kind of assessment has been conducted informally and qualitatively. However, the net result is controversy, with investigators judging the bias small and critics judging it large. Further, 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 non-differential misclassification of disease proposed by Horwitz and Feinstein[28] to explain associations between early exogenous estrogen preparations and endometrial cancer. When proper sensitivity analyses were conducted on this bias, only a negligible proportion of those associations were explained by bias.[28–30]

Epidemiologic applications of quantitative methods with long history in the decision sciences have become accessible for quantifying uncertainties about multiple sources of systematic error in a probabilistic manner.[26,31–33] 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 in assessing selection bias in a study of topical coal tar therapy and skin cancer among severe psoriasis patients;[31] exposure misclassification and selection bias in a study of phenylpropanolamine use and stroke;[32] and selection bias, confounder misclassification, and unmeasured confounding in a study of less than definitive therapy and breast cancer mortality,[26] as well as in other clinical and non-clinical applications.[33–35] 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. 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 not only supplement, but can 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.[36]

Currently available solutions

Best practices for questionnaire design

Designing a questionnaire for collecting epidemiologic data requires careful planning and pretesting before fielding the study, and requires validation of response during the analysis phase to make sure the data being collected are as valid as possible for addressing the study hypothesis. The following steps should be considered during the design and analysis stages of a study requiring data collection via a questionnaire:

- Use validated instruments or validated questions whenever possible.
- Strive for a fifth grade literacy level if you must develop new survey questions to be used for a general population.[37] Use cognitive testing to assess respondent comprehension of new questions.
- Evaluate the accuracy of respondents' answers by comparing them to a truly accurate comparison source (i.e., gold standard) whenever possible so that sensitivity and specificity can be calculated for use in bias analyses.
- Compute the percent agreement and kappa to test the reliability of self-report. For example, when evaluating medication use, compare with pill counts, chemical markers inserted into the pills, electronic monitoring caps, or pharmacy dispensing databases.
- Assess validity and reliability on a subset of the respondent population.

As in history taking during a clinical visit,[38] epidemiologic research using questionnaires often relies on asking respondents to recall events or exposures that occurred in the past, with recall intervals spanning days to several years. To appreciate the accuracy of data derived by recollection, one must understand the response process in general and the organization of memory, a key element of the response process.

Measurement error for survey data depends on the adequacy of the response process, which is made up of four key respondent tasks: (i) question comprehension and interpretation; (ii) search for and retrieval of information to construct an answer to the question; (iii) judgment to discern the completeness and relevance of memory for formulating a response; and (iv) development of the response based on retrieved memories.[39–42] If survey instrument developers pay too little attention to the first two key tasks, this can result in questions too vague or complex for respondents to marshal retrieval processes appropriately. We will not go into depth on the theory of survey response or cognitive process underlying retrieval and questionnaire response but refer readers to a text by Tourangeau and colleagues[42] for an enlightening discussion of these topics.

Most pharmacoepidemiologic research requires assessing the timing between when an exposure occurs and when an outcome is observed, which

may range from several hours to many years. Thus, questionnaires used in pharmacoepidemiology typically include one or more different types of temporal questions[42] such as:

- time of occurrence, which requires respondents to provide a date when an event occurred such as when were they 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?"; and
- 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 6 months?"

An example best illustrates the theory of Tourangeau and colleagues[42] on how respondents use a cyclic process of recalling details about a particular event. As new information is recalled, this new information helps shape the memory and adds details to describe the event in question: "When was your major depression first diagnosed?" The respondent may use the following process to provide the correct date, namely January 2008.

The recall process begins with the respondent being uncertain whether the depression was diagnosed in 2007 or 2008. To work towards identifying the correct year, the respondent recalls that the depression was the result of his losing 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-2007, just as their children finished the school year. So, in 2007 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 2008? It was January 2008 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 diagrammed in Figure 41.6.

As illustrated in Figure 41.6, landmark events probably serve as the primary organizational units of autobiographical knowledge and, as such, anchor

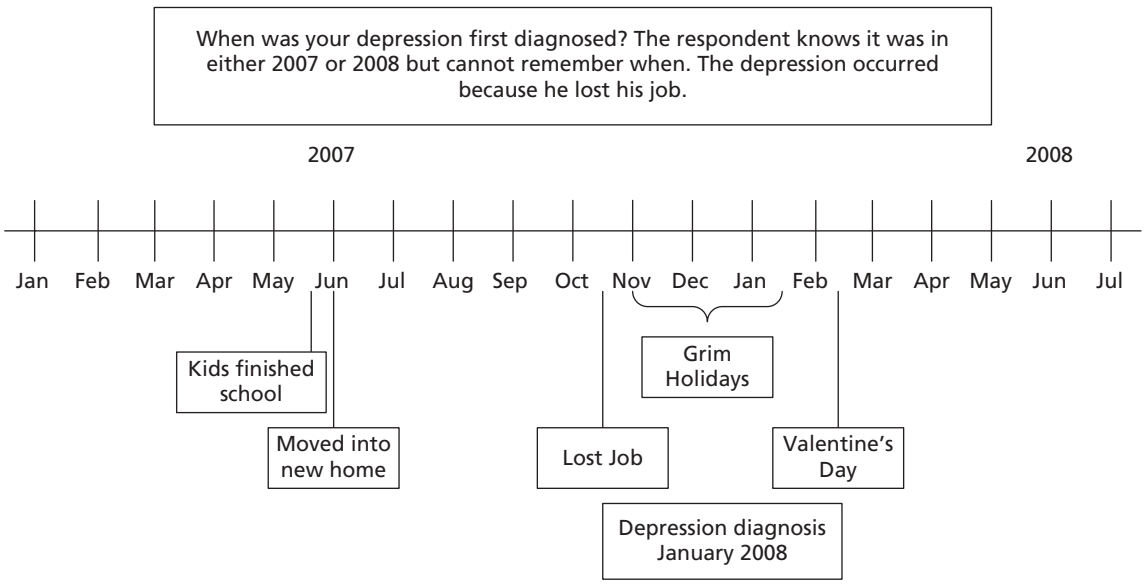


Figure 41.6 Recall schematic for showing how date of depression diagnosis was determined.

information retrieval.[43] 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 was first diagnosed. An important caveat is that the respondent described above 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.

In contrast, 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.[44,45] 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, respondents more frequently choose answers at the top of the list rather than those at the bottom, to minimize effort. Respondents with lower cognitive skills and less education, where discerning the best possible response poses a challenge, 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.

The discussion above 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. Understanding the measurement error associated with key variables critical to the analysis can be assessed using several different modeling approaches, which Biemer discusses in more detail.[46]

Conducting validation studies to assess self-reported data

Exposure confirmation performed as part of etiologic studies is often only partial verification for two reasons. First, the comparison data source may be an alloyed gold standard, which means the rate calculated is a measure of agreement, not a measure of validity. Second, and more commonly, verification studies using a gold or an alloyed gold standard can assess only one of the two validity measures, either sensitivity or specificity.

Methodologic studies that use alternative data sources such as prospectively collected data or databases of dispensed drugs can measure both sensitivity and specificity, if one assumes that the prescription database is a gold standard. Lower sensitivity is often more of a concern than is lower specificity, depending on the data source used for the study. Drug exposures or diseases that are under-reported on questionnaires or are missing due to incomplete claims processing in a record-linked database—that is, data sources with low sensitivity—cannot be rigorously evaluated as risk factors for the association under investigation. 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 rare, a small degree of misclassification when using administrative claims data where the case definition uses the ICD-9-CM code 695.1 will include several skin problems other than Stevens–Johnson (i.e., the false-positive rate would be high).[47]

Besides the need for completeness on the individual level, information from all persons who are covered by the health plan from which the database is generated must appear in the database. 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

The early work on evaluating the completeness of self-reported medication data used paper medical records for comparison, but one needs to

understand the availability and accuracy of medical records to determine whether they are adequate for this purpose. 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 also on the health-care provider's attention in recording the information, and on the availability of the medical record for review. If the medical record cannot be retrieved because the health-care provider could not be identified, had retired, or the record was destroyed or lost, the events cannot be verified.

Even if the outpatient and inpatient medical record is available, it may be incomplete for medications prescribed. Three studies have shown that the inpatient medical records are often incomplete for documenting medications patients are using in the outpatient setting. In comparing inpatient medical records with patient self-report combined with pharmacy dispensing,[48] Lau and colleagues found that the inpatient medical records typically omitted medications that patients used rather than include medications not mentioned by the patient or dispensed according to pharmacy records. Similarly, Strom *et al.* noted that drugs prescribed in the outpatient setting suspected of commonly causing Stevens–Johnson syndrome are often not documented in the inpatient chart.[49] Guess *et al.* reported similarly poor completeness when assessing whether the discharge and autopsy reports for patients with fatal upper GI bleeding or perforation indicated use of an NSAID that may have contributed to the fatal GI bleed.[50]

Completeness of medications in outpatient records depended on the number of drugs the patient was taking and the type of medication prescribed. The fewer the drugs the patient used, the more likely the drugs were listed in the chart.[51,52] Christensen[53] and colleagues reported excellent documentation of antihypertensives in the medical chart. Similarly, West *et al.*[54] reported very good entry of NSAIDs in outpatient medical records even though the patients had only been dispensed one fill of one NSAID in their entire pharmacy claims file. However, medications considered to be more sensitive, that is psychotropic medications such as benzodiazepines, tended to be omitted more frequently than non-psychotropic medications.[55]

In summary, the medical record does not document all medications prescribed for individuals. Record completeness is likely to vary by type of drug, type of chart (outpatient versus inpatient), and the number of drugs prescribed in a given period. This diminishes the usefulness of medical records for verifying self-reported drug exposure.

Medical records are often used to validate self-reported diagnosis data, but the diagnosis documentation in the medical record may be incomplete as well. With conversion from paper to electronic medical records (EMRs), more studies are evaluating the completeness for the EMRs rather than paper medical records. In 2008, three studies examined the completeness of paper obstetric records. One study, of bleeding, showed sensitivity ranging from 28.3% to 100% with better documentation with severe bleeding.[56] Another study evaluating documentation of obstetric care noted that the head to body delivery interval was recorded in only 58.2% of shoulder dystocia deliveries, even though it is a key measure in these types of deliveries.[57] Similarly, in a study of 190 operations performed over a 1-week period in one hospital, either the patient's name, preoperative diagnosis, type of operation, or postoperation instructions was missing in a small percent of charts, with only 51.6% of operative notes having complete documentation.[58] A 2007 obstetric publication reported better documentation of obstetric care by midwives than by attending physicians or residents.[59]

Two evaluations of EMR completeness noted that the patient problem list was often incomplete[60] and the anesthetic record that is supposed to document allergies, intravenous drug access, electrocardiogram rhythm, and issues related to ventilation often were missing these critical attributes.[61] Driscoll and colleagues attributed these omissions to incompatibility between the organization of the record and clinician workflow. We are likely to see more publications evaluating the completeness of the EMR with increasing use of this newer technology and as the documentation of clinical perform-

ance and care quality becomes a critical factor in health-care delivery.[62]

Self-reported drug data from *ad hoc* questionnaire studies

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.[63–71] Researchers can facilitate recall and reporting of medication use by indication-specific questions, a drug photo prompt, list of drug names, or a calendar to record life events.[70–74] In general, greater inaccuracies have been noted as more time elapsed between occurrence of exposure and its subsequent reporting;[64,68,70] this tendency was especially true for over-the-counter NSAID use in contrast with prescription NSAID use for recall over a 2-month period.[75] Accuracy of self-reporting varies by medication, with chronically used medications (especially those with more refills) recalled more often than acute exposures, first and most recent brands in a class recalled more frequently than other medications in the class, multiple medications in one class recalled more frequently than single medication exposure, and salient exposures (those that prompted study initiation) more accurately recalled than common and less disconcerting exposures.[63,64,69,73,76–79] 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.[80] Discrepancies are due to both under-reporting (e.g., respondent forgot medication was taken) and under-documenting (e.g., physician was unaware of medication use or did not record patient's use in chart),[52–54,65,76,77,79–82] and differed by therapeutic class.[80,81,83–90] When comparing self-reported data to multiple sources (e.g., medical records and pharmacy dispensing), verification for self-reported use was higher than with comparison to a single source.[91]

Influence of questionnaire design

As reported in a recent systematic review, several factors affect the accuracy of medication exposure reported via questionnaire.[92] The type of question asked influences how well respondents answer medication questions, illustrated by a study seeking use of analgesics in the past week.[93] Design also influences the completeness of self-reported psychoactive medication use.[94] Medication-specific or indication-specific questions can identify most medications in current use, and a general medication question "Have you taken any other medications?" failed to identify all of the medications respondents were currently taking.[79] 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.[11] The addition of indication-specific questions to open-ended questions also adds incremental affirmative responses concerning exposures. Finally, 20–35% of respondents reported drug exposure only when asked medication (name)-specific questions.[11] Similar findings were recently reported for self-reported medication use in a university population.[92]

Response order may affect recall, as noted with malaria medications when respondents had more than one episode of malaria.[95] Medications listed earlier tended to be selected more frequently than those listed later—a finding that may be related to satisficing as discussed earlier.[44]

A comparison of self-report for current and recent medication use (within the past 2 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%).[96] 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 finding might have occurred: that respondents

were unable to recall all medications when this self-reported information was compared to 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.[79] The questionnaire had indication-specific questions first, for example medications used for hypertension, diabetes, etc., 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 the results on sensitivity of recall, indication-specific questions appear to invoke better recall accuracy. However, to adequately address the issue of question structure, the questionnaire might have been designed to query medications using an open-ended question before asking 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. No differences in recall accuracy were noted by gender.[70,96] Inconsistent results were noted with ever having used an antidepressant for evaluation of age, household income, and education as predictors of recall accuracy for reporting.[97] Racial and socioeconomic differences in reporting were noted with oral contraceptives use, with whites having better agreement than non-whites and private paying users having better agreement than those receiving public health-care funds.[64] Small variations in recall accuracy were noted for any past estrogen use by ethnicity (Japanese vs. non-Japanese ancestry) and education, with more educated women having poorer recall than those without a college education.[68] A study of NSAID use noted a similar finding for education.[78] Non-smokers had better recall of ever/never use of estrogens than did smokers in one study.[68] Another study found no relationship between recall accuracy for past NSAID or estrogen use and cigarette smoking or current alcohol use.[70] A study of medication use during pregnancy reported better recall in mothers with higher educational attainment and poorer pregnancy outcome (low birth weight, gestational age, or Apgar score)[82] whereas other authors found that factors such as maternal age, marital and employment status, and pregnancy outcome did not influence the reporting of pregnancy medication exposures.[76,77,81]

Medication-specific questions substantially increased reporting for certain subgroups, including 25–44 year olds, males, African Americans, and those with 8 or more years of education.[11,79,94] Age affected recall accuracy for hormone shots,[90] NSAIDs,[50,70] and other medications,[96] with younger respondents having better recall accuracy. However, this finding did not hold true for oral estrogen[68] or oral contraceptive use.[64] Study design may explain the different results noted; the two studies that reported an age effect were methodologic studies evaluating recall accuracy[70,96] whereas the two that reported no age effects[64,68] were etiologic studies that reported verification of drug use as a measure of exposure misclassification for the association under study.

With regard to predictors of recall accuracy, factors such as questionnaire design, use of memory aids, recall period, extent of past drug use, age, and education sometimes influence how well respondents remember past drug use, the effect often seeming to vary by therapeutic class. Behavioral characteristics such as smoking and alcohol use were rarely evaluated as predictors of accuracy, and inconsistent findings were noted in the two studies that reported the results of their evaluation. Because of the paucity of information on predictors of recall, further research in this area is warranted.

Conclusions

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 *ad hoc* case–control studies. Researchers are using more 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 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 hospitalization data from *ad hoc* studies

Accuracy

Just as recall accuracy of past medication use varies by the type of drug, the ability to remember disease conditions varies by disease. The best reporting has been noted with conditions that are specific and familiar, such as diabetes mellitus,[87,98–104] hypertension,[87,99,101,102,105] asthma,[98,100,101] and cancers such as breast, lung, large bowel, and prostate.[102,105–107] However, assessing reporting accuracy is more difficult for common, 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.

In studies comparing self-reports to clinical evaluation, depending upon the type of condition, both under- and over-reporting have been found.[100,101] Studies using medical records to assess recall accuracy for common ailments typically found poor agreement, where under-reporting was often the major cause of the disagreement in some studies.[87,98,102,104] Over-reporting occurred as well, especially for conditions where the diagnostic criteria are less explicit.[108] Comparing self-reported symptom and quality of life information at two different time periods also shows both over- and underestimation as noted with pain recall preoperatively and 3 months after total knee arthroplasty,[109] 1.5 to 3.8 years after hip replacement,[110] and 5 to 10 years after back pain.[111] Subjects also recalled better walking and function than they actually had prior to the hip replacement.[110] Thus, recollection of symptoms and function are not accurate and the direction of the error can lead to both over- and underestimation.

Two studies assessed the recall accuracy for mental illnesses, comparing interview to clinical evaluation.[100,101] The results indicated poor agreement between the two data sources, with under-reporting as the primary reason for poor agreement. It is unclear from these studies whether the reason for under-reporting was the respondent's unwillingness to admit to mental illness or whether the conditions were actually under-diagnosed.

Three studies evaluated reporting of cataracts, two assessing presence of cataract by clinical exam[100,112] and the third using medical record review for comparison.[99] Agreement was best with the medical record comparison,[99] whereas the studies that used clinical assessments typically reported poor agreement. Similar to the evaluation of mental illnesses, the question remains, could the under-reporting be due to patients being unwilling to divulge their diagnosis or, perhaps, are they unaware of their diagnosis in the first place?

Fractures were evaluated in four studies, all of which used medical records for comparison. Three studies indicated good agreement, although the one methodologic study of fracture incidence indicated a slight tendency for over-reporting of hand, finger, rib, or facial fractures.[113]

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 1 year was recalled by 90%.[72] Similarly for menopause, 45% of women were able to report their exact age at natural menopause and 75.5% reported age within 1 year. The percentage agreements for surgical menopause were 55.6% and 83.4%, respectively, for exact age and age within 1 year. Recall lengths were 7.6 years and 10.6 years for natural and surgical menopause, respectively. 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.[114]

Poor agreement with cardiovascular conditions is due to both over- and under-reporting, depending on the data source used for comparison.[87,99,101,102,104,105,107,115–117] Some studies have noted accurate recall of cardiovascular conditions,[101,102] though it has been suggested that their results were due to under-reporting of disease.[87] In most instances of recall error, many who had incorrectly reported 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.[107,115–117]

Influences on accuracy

Several factors influence reporting of a medical condition during an interview, including the type of condition and the subject's understanding of the problem. Reporting also depends on the respondent's willingness to divulge the information. Conditions such as sexually transmitted diseases and mental disorders may not be reported because the respondent is embarrassed to discuss them with the interviewer or worries about the confidentiality of self-administered questionnaires.[100,118] As a result, conditions considered sensitive are likely to be under-reported when ascertained by self-report.

Factors influencing accuracy of past diagnoses and hospitalizations include the number of physician services for that condition and the recency of services.[119–124] 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.[119,120,122] 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 better reported than those with little or no impact on lifestyle. More patients with current restrictions on food or beverage due to medical problems reported chronic conditions that were confirmed in medical records than those without these restrictions.[119] Similarly, those who had restrictions on work or housework reported their chronic conditions more often than those who did not have these restrictions.[119] 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 those occurring in the first 6 weeks of pregnancy.

Perhaps as a result of the emotional stress, lifestyle changes, and potential financial strain, hospitalizations tend to be reported accurately.[123] Only a 9% under-reporting of hospitalizations occurred where surgery was performed, compared to 16% of patients without a surgical procedure. Under-reporting in those with only a 1-day hospital stay was 28% compared with 11% for 2 to 4-day stays, and approximately 6% for stays lasting 5 or more days.

Researchers also agree that respondents remember the type of surgery accurately.[90,124–126] Recall accuracy was very good for hysterectomy and appendectomy,[84,98,102] most likely because these surgeries are both salient and familiar to respondents. Cholecystectomy[102] and oophorectomy[84] were not as well recalled and were subject to some over-reporting. However, over-reporting may have been due to the potential incompleteness of the medical

records used for comparison.[84] For induced abortions, marginal agreement occurred, as noted by records from a managed care organization: 19% of women under-reported their abortion history, 35% over-reported abortions, and 46% reported accurately according to their medical record.[127] More generally, 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.[125] Recall of surgery date (±1 year) was correct for 87.5% of patients interviewed.

Patient demographics and influences on accuracy

The influence of demographic characteristics on reporting of chronic illnesses has been thoroughly evaluated, although the results are conflicting. The most consistent finding is that recall accuracy decreases with age,[87,90,104,106,112] although this may be confounded by recall interval, or cohort (generational) effects. Whether gender influences recall accuracy is uncertain. Men have been found to report better than women, independent of age,[98] whereas conflicting evidence found that women reported better than men,[100] especially in older age groups.[119] Further studies indicate that gender and age differences depended upon the disease under investigation,[100] with women over-reporting malignancies and men over-reporting stroke.[104] No differences was found for reporting of hospitalizations by age or gender.[123]

Reporting of illnesses, procedures, and hospitalizations was better among whites than non-whites,[90,98,100,119,123,127] but the number of non-whites in studies was relatively small. Reporting by educational level was equivocal; one study showed no difference,[120] while another study indicated better recall for those with less education,[119] and others suggested more accurate responses for those with a college education.[104,106,113,127] Reporting was more complete for self-respondents compared to proxy respondents,[98,100,112,120] including reporting for hospitalizations, where under-reporting was estimated at 7% for self-respondents and 14% for proxies.[123] For self-respondents, those with a poor or fair current health status reported conditions more completely than those with good to excellent health status.[119]

Questionnaire design also influences validity of disease and hospitalization data obtained by self-report. Providing respondents with a checklist of reasons for visiting the doctor improves recall of all medical visits.[128] 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.[129] However, longer questions could increase the cost of the research and could needlessly tire the respondents, leading to satisficing.

In summary, whether a person reports an illness during an interview appears to be related to age and the 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.[100,119,123]

Example

As indicated previously, accuracy of *ad hoc* 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 41.1.

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.[70] While a single dispensing was reported

Table 41.1 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[70]	2–3 years 7–11 years	Telephone interviews n = 319	Non-steroidal anti-inflammatory 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:
						±6 months / ±1 year / ±2 years First use: 20 / 28 / 51 Last use: 17 / 24 / 42 Duration: 67 / 71 / 80
Smith 1999[89]	Current use	Personal interview and medication inventory n = 55 users	Aspirin	None	Serum levels	0.16 (0.0–0.32)

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. Smith *et al.* compared current use of aspirin (among other medications) as stated in a personal interview and medication inventory to blood serum levels.[89] They found minimal agreement ($\kappa = 0.16$, 95% CI: 0.0–0.32) between statement of use with inventory compared with serum levels of medication.

Continuing the NSAID 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 studies for diagnoses (Table 41.2). 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. Ambegaonkar and colleagues compared the new, 11-question Gastrointestinal Toxicity Survey to the six-question Stanford Calculator of Risk for Events (SCORE) for accuracy in identifying patients at high risk for NSAID-associated GI events:[130] the correlation between the two questionnaires was 0.96 ($p < 0.001$). Using two types of regression analysis, the agreement of the two assessments across four risk categories was 79.8% and 88.8%, respectively. Fourrier-Reglat, *et al.* compared reported medical data from patient and prescriber self-administered questionnaires.[131] MI 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. Using the prescriber data as the gold standard, patient reports of MI provided moderately complete data (sensitivity: 77.7%, specificity: 99.6%, PPV: 77.1%, NPV: 99.6%); reports of upper GI bleeding by patients were not confirmed by the prescriber reports (sensitivity: 44.6%, PPV: 10.4%).

For an evaluation of NSAID-induced GI complications, Singh and colleagues noted that prior validation had been conducted for the Stanford Health Assessment Questionnaire.[132] This validation took place in multiple steps. First, patients were queried to determine whether they understood the symptoms as described using lay language. When there were misunderstandings, the questions were modified accordingly. Patients were then given the questionnaires multiple times, interviews were conducted, and the results were compared with physician and hospital records to determine patient recall and accuracy. To evaluate the validity of a questionnaire assessing potential adverse drug events associated with NSAID use translated from English into Thai, Jarernsiripornkul *et al.* also utilized a multistage process.[133] Five health professionals determined the clarity and meaningfulness of symptoms in the questionnaire and scored the consistency of ratings. Questions were changed based on the scoring, and the new questionnaire was piloted among patients similar to those who would be included in the study. Interviews were conducted with this pilot group to determine patient understanding. Medical doses and pictures of the products included in the questions were added during the process to facilitate recognition.

These examples demonstrate the variation in methods used to 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.

Validity of pharmacoepidemiologic drug and diagnosis data from computerized databases containing administrative or electronic medical record data

In addition to conducting *ad hoc* studies to evaluate drug–disease associations, a variety of computerized, administrative databases are available for pharmacoepidemiologic research, the structure, strengths, and limitations of which are reviewed in Chapters 11–18. One major advantage of using such databases for pharmacoepidemiologic research is the comparative validity of the drug data in lieu

Table 41.2 Validation of MI or GI outcomes in patients with NSAIDs in questionnaire data

Author	Questionnaire and sample size	Study question	Comparison data source	Conditions	Findings
Ambegaonkar, 2004[130]	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 4 risk categories and an 88.8% agreement when the 2 highest risk categories were collapsed into a single category The multinomial logistic regression (MNL) analysis showed agreement of 75.8% for 4 risk categories and an agreement of 86.8% for 3 risk categories For both methods, disagreement was equally distributed among over-prediction and under-prediction of risk levels by the GITS (NI) questionnaire relative to the SCORE questionnaire; in the case of 4 risk categories, disagreement by 2 risk levels was limited to 0.6% and 1.5% for the FGLS and MNL regression methods, respectively

Fourrier-Reglat 2010[131]	CADEUS cohort (French national cohort study of traditional NSAIDs and Cox-2 users conducted between September 2003 and August 2004 in France that employed self-administered questionnaires to obtain medical data from patients and their prescribers) n = 18530 pairs of patients and prescribers	To compare patients and prescribers reported medical data	Prescribers report as gold standard	Previous medical history: MI: k = 0.75 (95%CI: 0.71–0.80) Sensitivity: 77.7% Specificity: 99.6% PPV: 74.1% NPV: 99.6% Upper digestive hemorrhage: κ = 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 (κ = 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[132]	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 under-reporting by patients, those events that are severe enough to require hospitalization are also ascertained by record review of all hospitalizations.

of questionnaire data, where recall bias is always a concern, as previously described.

In general, the databases differ widely on many factors, such as size (e.g., from several hundred thousand to several million covered lives), number of plans included, the type of health services provided and therefore available for analysis (e.g., prescriptions, mental health benefits, etc.), whether out-of-plan claims are included in the main database or resident in other databases, and the timeliness of the data (e.g., the lag for prescriptions is typically in weeks whereas that for outpatient visits may be 6 or more months). The databases also differ on the number of available demographic variables: all have age and sex, but few have race, occupation, or a measure of health status.[134] Because the plans 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).

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 United States, typically through the place of employment. If the employer changes plans, which may be done 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. Thus, the continual enrollment and disenrollment of plan members hinders the opportunity for longitudinal analyses. It is unclear whether and how enrollment and longitudinal follow-up capabilities are expected to change with the advent of the 2010 Patient Protection and Affordable Care Act.[135]

Along these lines, 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 and/or events of interest 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.[136] For example, completeness of the drug data might vary by income level if persons with higher incomes and drug copayments 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 from the database.

Best practices for validation studies in administrative or medical record databases

For the data in an administrative database 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 medical records, administrative or billing records, pharmacy dispensings, or procedure logs. Choice of an appropriate comparator varies by study question, information utilized from the database and comparator, and availability of other data sources. The study investigator must be aware of the limitations of both the administrative database and the chosen comparison dataset. For instance, over-the-counter medications are unlikely to be available for study in either administrative claims or pharmacy dispensing records. The chosen comparator should provide sufficient data to validate both the exposure and outcome algorithms used for the study and to evaluate the completeness and accuracy of the chosen cohort. A variable that provides exact linkage between the datasets, such as a medical record number, should be available so that exact algorithms can be evaluated for accuracy within a subset of known study patients. For example, if a single claim contains six diagnosis codes and 6 months of claims were used to determine outcomes in patients, then all six diagnosis codes for all claims across the 6-month study time must be available in a comparison dataset to establish validity of the algorithm used for 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 predic-

tive value, sensitivity, and specificity combined provide a complete picture of the agreement between the two datasets.

The following is a broad overview of how to conduct a validation study in administrative 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, abstract variables needed to determine cohort selection, exposure, outcome, and other variables for validation. Calculate measures of agreement between the two datasets. Finally, consider strengths and limitations of the two datasets to ascertain validity and completeness of the administrative database to answer the study question.

In addition to validation and completeness described above, analyses conducted to evaluate the usefulness of administrative databases for observational studies include assessing the following three factors: consistency between data files within the same system, surrogate markers of disease, and time-sequenced relationships, such as a diagnostic procedure preceding a surgery.[137]

Drug data in administrative or medical record databases

Prescriptions can be written, dispensed, but not picked up (unclaimed) by patients. Unclaimed prescriptions, estimated to occur for approximately 2% of all prescriptions, present an adherence issue in administrative data.[138] For every 1000 new prescriptions, an average of 16.5 are unclaimed.[139] Anti-inflammatory and anti-infective drugs tend to be the therapeutic classes most often unclaimed.[139–142] Two-thirds of unclaimed prescriptions were for new prescriptions,[140] and a similar proportion tended to be non-essential medications.[143] Many unclaimed prescriptions were telephoned in,[138,140] and the most frequently cited reasons for not picking up prescriptions were that the patients determined that they did not need the medication or they forgot to pick it up.[138,141] However, cost and having a similar medication at home were also often cited.[138,141,143]

One might ask how unclaimed prescriptions might affect the validity and completeness of pharmacy data. Many individuals have some type of pharmacy benefits plan where reimbursement for medication costs goes through a third-party payer. Entry into the reimbursement software is predicated on dispensing of the drug. However, a drug that is dispensed but is not claimed should be returned to stock and the appropriate adjustment made to the patient's pharmacy benefits plan—it is insurance fraud if this were not to happen. Unfortunately, we do not know whether all such insurance adjustments have been made so, as researchers, 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 whatever reason.

Drug data in administrative databases are often not validated. Administrative data cannot address adherence and drug ingestion, and over-the-counter medications are not typically included. Thus, although most researchers are comfortable that claims for a drug are an accurate and complete representation of exposure to that drug, this assumption may not be accurate and should be tested when using a new drug exposure or database. Similarly, sensitivity analyses should be performed to determine the susceptibility of the results to possible misclassification.

Diagnoses and hospitalizations in administrative databases

Unlike the drug data in administrative databases, 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 persons 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,[144] 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 record-linked databases are often coded using the International Classification of Disease (ICD) coding system. Poorly defined diseases are difficult to code using the ICD system and there is no way to indicate that an ICD code is coded for "rule-out" purposes. How health-care plans deal with "rule-out" diagnoses is unclear, that is, are they 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.[145] In addition, reimbursement standards and patient insurance coverage limitations may influence the selection of ICD codes for billing purposes.[146] 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 would be less likely to occur in staff/group model HMOs such as Group Health Cooperative or Kaiser Permanente. Lastly, ICD version changes may produce systematic errors.

Continuing with the NSAID example, we conducted another literature scan of published studies validating MI or GI bleeding outcomes with use of NSAIDs in administrative databases; these studies are summarized in Table 41.3. Administrative data are often compared with medical records in a validation study. Most of these studies provide only positive predictive value (PPV) that indicates whether the coding scheme is accurately classifying observed measures as 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,[148] Saskatchewan Health,[149] and the HealthCore Integrated Research Database.[150] In all of these databases, this ICD-9-CM code had substantial or nearly perfect ability to classify MI in medical records as MI. In the Quebec computerized health database, ICD-9 code 410.xx had a PPV of 0.96 (95% CI: 0.94–0.98).[148,151] In Saskatchewan Health, the PPV for MI as measured by presence of ICD-9-CM code 410.xx compared with medical records was 0.95 (95% CI: 0.91–0.98). Both the overall PPV for ICD-9-CM 410.xx to measure MI and the PPV for MI among patients taking NSAIDs were denoted in the HealthCore Integrated Research Database. Among all of the patients with a code for MI, the PPV was 88.4% (95% CI: 83.2–92.5%). Among patients taking 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 highlights 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 results.

Other ICD-9-CM codes used for possible detection of MI have shown poor ability to classify MI. Use of ICD-9-CM 411.xx to detect acute MI in Saskatchewan Health data[149] yielded a PPV of only 0.09 (95% CI: 0.07–0.11). Other measures of MI have been assessed to varying degrees in other administrative records. A substantial proportion of cases meeting 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 for Puget Sound[152] (sensitivity 86.5%, specificity 85.4%) and the General Practice Research Database[153] (sensitivity 89.3%), with substantial agreement between the administrative and medical records.

Measurement of GI bleeding is more varied across databases and several algorithms using different combinations of ICD-9-CM and CPT codes have been used to determine event occurrence. Validation of GI events in Veteran's Affairs (VA) administrative data was conducted in an iterative manner.[154] Sensitivity and specificity were higher when using only ICD-9 or CPT codes, but the PPV

Table 41.3 Validation of MI and GI events in studies using administrative data to evaluate harms of NSAID exposure

Author	Dataset	Study sample size	Study aim	Comparison data source	Conditions	Findings
Abraham, 2006[154]	VA	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)	To validate Veterans Affairs (VA) administrative data for the diagnosis of NSAID-related upper gastrointestinal events (UGIE) and to develop a diagnostic algorithm	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: Sensitivity: 100% Specificity: 96% PPV: 27% NPV: 100% ICD-9 and CPT for UGIE: Sensitivity: 82% Specificity: 100% PPV: 51% NPV: 99% ICD-9 and CPT algorithm for UGIE: Sensitivity: 66% Specificity: 88% PPV: 67% NPV: 88% Algorithm validated in additional 44 patients, PPV among NSAID users: 80%
Brophy, 2007[148]	Computerized health databases of Quebec, Canada	n = 234 MI survivors	To determine whether a history of MI modified the risk of acute MI associated with the use of various NSAIDs	Previous validation of MI claims;[149] 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 41.3 (*Continued*)

Author	Dataset	Study sample size	Study aim	Comparison data source	Conditions	Findings
Castellsague, 2009[155]	Saskatchewan Health	Specific codes: n = 38 (10% sample) Non-specific codes: n = 742 (all potential cases)	To estimate the risk of upper gastrointestinal complications associated with use of cyclooxygenase-2 (COX-2) selective (celecoxib and rofecoxib) and individual non-selective non-steroidal anti-inflammatory drugs compared with non-use of these drugs	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 non-specific codes = 68%. This study: Specific PPV = 92% Non-specific code PPV (ranged across codes) = 60% for unspecified hemorrhage, 4% for hemorrhage of rectum/anus
Curtis, 2008[147]	Medicare	Number not specified.	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	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[158]	GPRD	n = 96	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	Medical records	Upper GI events: ICD-10 codes K25–K29 NSAIDS: any prescription in GPRD	PPV = (95/96) = 99.0%
Varas-Lorenzo, 2009[149]	Saskatchewan Health	n = 200	To evaluate risk of fatal and non-fatal acute MI with NSAID use	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 non-fatal) 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 41.3 (*Continued*)

Author	Dataset	Study sample size	Study aim	Comparison data source	Conditions	Findings
Wahl, 2010[150]	HealthCore Integrated Research Database	n = 200 charts per outcome	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	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%)

for determining a GI event increased with combined assessment of ICD-9 and CPT codes. While the sensitivity dropped from 100% when assessing only one ICD-9 code to 66% when using the combined diagnostic algorithm (and, similarly, the specificity dropped from 96% to 88%), the PPV increased from 27% with one ICD-9 code to 67% with the combined diagnostic algorithm. Limiting further to only those patients using NSAIDS, the PPV increased to 80%. Assessment in Saskatchewan Health also included GI bleeding as one of many upper GI complications potentially associated with NSAIDs. The PPV for non-specific GI codes was 68% and for non-specific codes indicative of bleeding the PPV was 60%.[155] Another study assessing specific GI bleeding codes had a composite PPV close to 90%.[156] Upper GI events in the General Practice Research Database were assessed via ICD-10 codes, and 95 of the 96 charts reviewed indicated that a GI event had occurred, for a PPV of 99.0%.[157,158] 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.[150] Among all patients with an ICD-9 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%).

Other studies have been conducted assessing GI bleeding classifications in administrative data compared with medical records, independent of whether the patients were taking NSAIDs. In comparing claims from the HMO Research Network to medical records, only 23% of patients identified in claims had confirmed peptic ulcers and bleeding,[159] and only 28% of hospitalizations for upper GI perforation, ulcer, or bleeding were confirmed.[160] Autopsy and medical record data were used as the comparator for GI bleeding leading to death in one assessment of Saskatchewan Health data.[50] Among the cases meeting the defined criteria in the administrative data, 76.8% had confirmation of GI bleeding in their medical record or at autopsy. Diagnoses from the Scottish Morbidity Record were compared with original hospital records with a PPV of 46.6%.[161] Comparison of hospital medical records to the Tayside Medical Monitoring Unit (MEMO) data yielded a sensitivity of 68% for acute GI bleeding and 79% for perforation.[162] The specificity for both admission diagnoses was 98%.

Two studies were identified that are not discussed above or included in Table 41.2. One study was conducted using data from the General Practice Research Database and assessed MI.[163] The other study was conducted using data from The Health Improvement Network (THIN) and assessed upper GI complications, including GI bleeding.[164] Although both studies assessed the PPV of undisclosed READ codes plus investigator determination of outcomes from free text, the nature of their algorithms would be difficult for other researchers to use.

The variation seen in comparisons of GI bleeding in administrative records to that found in medical records may be due to the differences in algorithms used to determine GI bleeding. The variation may also be due 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 administrative databases will aid in determining whether GI bleeding can be adequately assessed in this type of data.

Conclusions

Validating the case definition developed for observational studies using administrative databases with original documents such as inpatient or outpatient medical records is a necessary 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, there is still a need for validation of drug exposures and disease diagnoses in other administrative databases or in instances where 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. Completeness is typically assessed for a particular component of a study, such as the effect of drug copayments on

pharmacy claims[165,166] or the availability of discharge letters in the General Practice Research Database.[153] A study published in 1984[128] indicated that pharmacy data from administrative databases were of high quality, and because claims are used for reimbursing pharmacy dispensings, this should continue to hold true today. We realize that adherence is an issue (see Chapter 42) and that not every dispensing indicates exposure, but we do not know the extent of unclaimed prescriptions and whether this might affect our research. Although administrative databases have greatly expanded our ability to do 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

This chapter describes the methodologic work that has been published on data quality issues about the conduct of pharmacoepidemiologic studies, whether the medication and diagnosis information arises from questionnaires, administrative claims, or EMRs. We also discuss how to minimize measurement error in epidemiologic studies, assess whether measurement errors occurred by validating important study variables, and evaluate the impact of these errors on the direction and magnitude of effect.

Methods for conducting pharmacoepidemiologic studies have shifted over the past 25 to 30 years from reliance on studies requiring *de novo* data collection from individuals, to extensive use of electronic data from either administrative health claims or electronic medical records, and in the near future, to studies using distributed data networks (see Chapter 30) and data from health information exchanges. Yet, *de novo* data collection will continue to be required to ascertain information on quality of life, patient-reported outcomes, and medications either not included in pharmacy dispensing files or not reliably entered into EMRs, such as herbal and over-the-counter medications.

In contrast with *de novo* data collection for pharmacoepidemiologic research, the availability, use, and richness of electronic data sources have increased exponentially. At least 25 years ago, researchers realized that administrative claims were quite useful for conducting pharmacoepidemiologic studies,[167,168] yet there were critics.[169] Although drug data from Medicaid claims were reliable—that is concordance existed between the drugs prescribed and dispensed—the concordance for diagnosis claims was not very good.[128]

With time, we found that careful and systematic evaluation of data accruing for administrative purposes could be used to study drug–disease relationships without spending the time and money to collect data *de novo*. We learned how to develop algorithms containing ICD-9 diagnosis, procedure, and external causes of injury codes that were used as billing codes in the administrative claims to identify individuals with certain diseases, for example, acute myocardial infarction.[169,170] Just the occurrence of the procedure code, without the results of the procedure, provided useful knowledge on whether the individual had the disease under study. Ultimately, researchers realized that any study using electronic data would require verification of diagnoses to ensure the validity of the case definition. This extra time and cost burden for verification was minimal compared to the time and budget required for full data collection. Record-linked databases with automated pharmacy files could also minimize measurement error for exposures because drug data would be identified by dispensing records, not by self-reports, with the potential for recall bias and exposure underascertainment.

The improved computer technology that resulted in faster processor speeds and increased storage capacity facilitated storage of health-care data in an electronic format, that is EMRs, 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 that require knowledge not only about whether a procedure or laboratory test was done, but also the results of these clinical events. The obvious advantage of access to electronic clinical data is less reliance on the need to confirm diagnoses using paper medical records, especially

when there is little or no paper copy backup to review.

Although clinical practice in the United States has been moving toward EMRs slowly, the Health Information Technology for Economic and Clinical Health (HITECH) provisions of the 2009 American Recovery and Reinvestment Act[171] established financial incentives for US providers to begin using EMRs starting in 2011. The increasing uptake of EMRs will lead to increased availability of more granular clinical data for pharmacoepidemiologic research. However, similar to the validation process undertaken for research using administrative claims, data from EMRs will have to undergo scrutiny as well. Initial evaluation of EMR data suggests great promise, but a data mapping and standardization of terminology and codes will be required to make these data, collected for clinical care, useful for research.[131]

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 data to EMR data and to data from health information exchanges, as both claims and EMR data are linked. The concern that Lessler[128] and colleagues resolved about the validity of medication data for administrative claims is now being raised for prescribing data from EMRs—it is clear when the patient starts a medication but the duration of use may not be adequately documented. Diagnosis data from EMRs do not carry the same level of concern for validity as claims data because they are being used to care for the patients, not for reimbursement; however, confirmation of the accuracy assumption still is needed. In the future, because EMR data will increasingly be used for research, we hope and expect to see studies validating EMR data.

References

1. Gordis L. Assuring the quality of questionnaire data in epidemiologic research. *Am J Epidemiol* 1979; **109**: 21–4.
2. Brody DS. An analysis of patient recall of their therapeutic regimens. *J Chronic Dis* 1980; **33**: 57–63.
3. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.
4. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol* 1993; **137**: 1251–8.
5. Rothman KJ, Greenland S. *Precision and Validity in Epidemiologic Studies. Modern Epidemiology*. Philadelphia, PA: Lippincott, Williams, and Wilkins, 1998, pp. 129–30.
6. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987; **126**: 161–9.
7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**(8476): 307–10.
8. Kelsey JL, Thompson WD, Evans AS. *Methods in Observational Epidemiology. Monographs in Epidemiology and Biostatistics*, Vol. 10. New York: Oxford University Press, 1986.
9. Rothman KJ, Greenland S. *Precision and Validity in Epidemiologic Studies. Modern Epidemiology*, 3rd edn. Philadelphia, PA: Lippincott, Williams & Wilkins, 2008, pp. 141–2.
10. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; **32**: 51–63.
11. Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.
12. 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**: 537–46.
13. Kristensen P. Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiology* 1992; **3**: 210–5.
14. Thomas DC. Re: "When will nondifferential misclassification of an exposure preserve the direction of a trend?" *Am J Epidemiol* 1995; **142**: 782–4.
15. Weinberg CR, Umbach DM, Greenland S. Weinberg, *et al.* Reply—Re: "When will nondifferential misclassification of an exposure preserve the direction of a trend?" *Am J Epidemiol* 1995; **142**: 784.
16. Jurek AM, Greenland S, Maldonado G. How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? *Int J Epidemiol* 2008; **37**: 382–5.

17. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol* 2005; **34**: 680–7.
18. 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**: 746–8.
19. Poole C. Exceptions to the rule about nondifferential misclassification. *Am J Epidemiol* 1985; **122**: 508.
20. Rodgers A, MacMahon S. Systematic underestimation of treatment effects as a result of diagnostic test inaccuracy: implications for the interpretation and design of thromboprophylaxis trials. *Thromb Haemost* 1995; **73**: 167–71.
21. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977; **105**: 488–95.
22. Greenland S, Kleinbaum DG. Correcting for misclassification in two-way tables and matched-pair studies. *Int J Epidemiol* 1983; **12**: 93–7.
23. Willett W. An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. *Stat Med* 1989; **8**: 1031–40; discussion 71–3.
24. Raphael K. Recall bias: a proposal for assessment and control. *Int J Epidemiol* 1987; **16**: 167–70.
25. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol* 1996; **25**: 1107–16.
26. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003; **14**: 451–8.
27. Rothman KJ, Greenland S. *Basic Methods for Sensitivity Analysis and External Adjustment. Modern Epidemiology*, 2nd edn. Philadelphia, PA: Lippincott, Williams, and Wilkins, 1998, pp. 343–57.
28. Horwitz RI, Feinstein AR. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. *N Engl J Med* 1978; **299**: 1089–94.
29. Hutchison GB, Rothman KJ. Correcting a bias? *N Engl J Med* 1978; **299**: 1129–30.
30. Greenland S. A mathematic analysis of the "epidemiologic necropsy". *Ann Epidemiol* 1991; **1**: 551–8.
31. Greenland S. Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment. *Risk Anal* 2001; **21**: 579–83.
32. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003; **14**: 459–66.
33. Phillips CV, LaPole LM. Quantifying errors without random sampling. *BMC Med Res Methodol* 2003; **3**: 9.
34. 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.
35. 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**: 384–92.
36. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Springer, 2009.
37. Biemer PP, Lyberg L. *Introduction to Survey Quality*. Hoboken, NJ: Wiley-Interscience, 2003.
38. Barsky AJ. Forgetting, fabricating, and telescoping: the instability of the medical history. *Arch Intern Med* 2002; **162**: 981–4.
39. 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**: 982–92.
40. Tourangeau R. Cognitive sciences and survey methods. In: Janine TB, Straf ML, Tanur JM, Tourangeau R, eds. *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*. Washington, DC: National Academy Press, 1984, pp. 73–100.
41. 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.
42. Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. Cambridge, MA: Cambridge University Press, 2000.
43. 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**: 383–406.
44. Krosnick JA. The threat of satisficing in surveys: the shortcuts respondents take in answering questions. *Survey Methods Newsletter* 2000; **20**: 4–8.
45. Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly* 1987; **51**: 201–19.
46. Biemer P. Measurement errors in sample surveys. In: Pfeffermann D, Rao CR, eds. *Handbook of Statistics—Sample Surveys: Design, Methods and Applications*. The Netherlands: North-Holland, 2009, pp. 281–316.

47. Strom BL, Carson JL, Halpern AC, Schinnar R, Snyder ES, Shaw M, *et al.* A population-based study of Stevens–Johnson syndrome. Incidence and antecedent drug exposures. *Arch Dermatol* 1991; **127**: 831–8.
48. 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**: 597–603.
49. Strom BL, Carson JL, Halpern AC, Schinnar R, Snyder ES, Stolley PD, *et al.* Using a claims database to investigate drug-induced Stevens–Johnson syndrome. *Stat Med* 1991; **10**: 565–76.
50. Guess HA, West R, Strand LM, Helston D, Lydick EG, Bergman U, *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**: 35–45.
51. 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.
52. Monson RA, Bond CA. The accuracy of the medical record as an index of outpatient drug therapy. *JAMA* 1978; **240**: 2182–4.
53. 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**: 99–104.
54. 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**: 165–71.
55. Buchsbaum DG, Boling P, Groh M. Residents' underdocumentation in elderly patients' records of prescriptions for benzodiazepine. *J Med Educ* 1987; **62**: 438–40.
56. 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**: 481–4.
57. Crofts JF, Bartlett C, Ellis D, Fox R, Draycott TJ. Documentation of simulated shoulder dystocia: accurate and complete? *BJOG* 2008; **115**: 1303–8.
58. Lefter LP, Walker SR, Dewhurst F, Turner RW. An audit of operative notes: facts and ways to improve. *ANZ J Surg* 2008; **78**: 800–2.
59. Haberman S, Rotas M, Perlman K, Feldman JG. Variations in compliance with documentation using computerized obstetric records. *Obstet Gynecol* 2007; **110**: 141–5.
60. Galanter WL, Hier DB, Jao C, Sarne D. Computerized physician order entry of medications and clinical decision support can improve problem list documentation compliance. *Int J Med Inform* 2010; **79**: 332–8.
61. Driscoll WD, Columbia MA, Peterfreund RA. An observational study of anesthesia record completeness using an anesthesia information management system. *Anesth Analg* 2007; **104**: 1454–61.
62. Spies TH, Mokkink HG, De Vries Robbe PF, Grol RP. Which data source in clinical performance assessment? A pilot study comparing self-recording with patient records and observation. *Int J Qual Health Care* 2004; **16**: 65–72.
63. Glass R, Johnson B, Vessey M. Accuracy of recall of histories of oral contraceptive use. *Br J Prev Soc Med* 1974; **28**: 273–5.
64. Stolley PD, Tonascia JA, Sartwell PE, Tockman MS, Tonascia S, Rutledge A, *et al.* Agreement rates between oral contraceptive users and prescribers in relation to drug use histories. *Am J Epidemiol* 1978; **107**: 226–35.
65. 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**: 45–50.
66. 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.
67. 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.
68. 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**: 815–25.
69. Nischan P, Ebeling K, Thomas DB, Hirsch U. Comparison of recalled and validated oral contraceptive histories. *Am J Epidemiol* 1993; **138**: 697–703.
70. 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**: 1103–12.
71. 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**: 1210–3.
72. Bean JA, Leeper JD, Wallace RB, Sherman BM, Jagger H. Variations in the reporting of menstrual histories. *Am J Epidemiol* 1979; **109**: 181–5.
73. Coulter A, Vessey M, McPherson K, Crossley B. The ability of women to recall their oral contraceptive histories. *Contraception* 1986; **33**: 127–37.
74. Beresford SA, Coker AL. Pictorially assisted recall of past hormone use in case-control studies. *Am J Epidemiol* 1989; **130**: 202–5.
75. 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**: 211–8.
76. Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformations. *Am J Public Health Nations Health* 1967; **57**: 2071–5.
77. 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**: 37–45.
78. West SL, Savitz DA, Koch G, Sheff KL, Strom BL, Guess HA, *et al.* Demographics, health behaviors, and past drug use as predictors of recall accuracy for previous prescription medication use. *J Clin Epidemiol* 1997; **50**: 975–80.
79. 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**: 273–7.
80. 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**: 78–80.
81. Mackenzie SG, Lippman A. An investigation of report bias in a case-control study of pregnancy outcome. *Am J Epidemiol* 1989; **129**: 65–75.
82. 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.
83. Hulka BS, Kupper LL, Cassel JC, Efird RL. Medication use and misuse: physician-patient discrepancies. *J Chronic Dis* 1975; **28**: 7–21.
84. Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 1982; **116**: 114–22.
85. 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**: 672–6.
86. Johnson RE, Vollmer WM. Comparing sources of drug data about the elderly. *J Am Geriatr Soc* 1991; **39**: 1079–84.
87. Kehoe R, Wu SY, Leske MC, Chylack LT, Jr. Comparing self-reported and physician-reported medical history. *Am J Epidemiol* 1994; **139**: 813–8.
88. 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**: 997–1002.
89. 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**: 143–6.
90. Clegg LX, Potosky AL, Harlan LC, Hankey BF, Hoffman RM, Stanford JL, *et al.* Comparison of self-reported initial treatment with medical records: results from the prostate cancer outcomes study. *Am J Epidemiol* 2001; **154**: 582–7.
91. de Jong PCM, Prevoo MLL, Zielhuis GA, Roeleveld N, Gabreels F. Accessibility and validity of data on medical drug use during pregnancy collected from various sources. *J Pharmacoepidemiol* 1991; **2**: 45–57.
92. Gama H, Correia S, Lunet N. Questionnaire design and the recall of pharmacological treatments: a systematic review. *Pharmacoepidemiol Drug Saf* 2009; **18**: 175–87.
93. 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**: 529–34.
94. Cottler LB, Robins LN. The effect of questionnaire design on reported prevalence of psychoactive medication. *NIDA Res Monogr* 1984; **55**: 231–7.
95. 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**: 175–81.
96. 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 Sci* 1991; **13**: 91–6.

97. Cotterchio M, Kreiger N, Darlington G, Steingart A. Comparison of self-reported and physician-reported antidepressant medication use. *Ann Epidemiol* 1999; **9**: 283–9.
98. 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**: 1207–13.
99. 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**: 1554–6.
100. Commission on Chronic Illness. *Chronic Illness in a Large City: The Baltimore Study. Chronic Illness in the United States*, Vol. IV. Cambridge, MA: Harvard University Press, 1957, pp. 297–328.
101. 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**: 181–91.
102. 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**: 188–93.
103. Midthjell K, Holmen J, Bjorndal 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 Community Health* 1992; **46**: 537–42.
104. 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**: 1407–17.
105. Colditz GA, Martin P, Stampfer MJ, Willett WC, Sampson L, Rosner B, *et al.* Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol* 1986; **123**: 894–900.
106. Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill 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**: 556–62.
107. 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**: 101–6.
108. 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**: 409–17.
109. Lingard EA, Wright EA, Sledge CB. Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. *J Bone Joint Surg Am* 2001; **83-A**: 1149–56.
110. Mancuso CA, Charlson ME. Does recollection error threaten the validity of cross-sectional studies of effectiveness? *Med Care* 1995; **33** (4 Suppl): AS77–88.
111. Dawson EG, Kanim LE, Sra P, Dorey FJ, Goldstein TB, Delamarter RB, *et al.* Low back pain recollection versus concurrent accounts: outcomes analysis. *Spine* (Phila Pa 1976) 2002; **27**: 984–93; discussion 94.
112. 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**: 1438–46.
113. Nevitt MC, Cummings SR, Browner WS, Seeley DG, Cauley JA, Vogt TM, *et al.* The accuracy of self-report of fractures in elderly women: evidence from a prospective study. *Am J Epidemiol* 1992; **135**: 490–9.
114. Colditz GA, Stampfer MJ, Willett WC, Stason WB, Rosner B, Hennekens CH, *et al.* Reproducibility and validity of self-reported menopausal status in a prospective cohort study. *Am J Epidemiol* 1987; **126**: 319–25.
115. Tretli S, Lund-Larsen PG, Foss OP. Reliability of questionnaire information on cardiovascular disease and diabetes: cardiovascular disease study in Finnmark county. *J Epidemiol Community Health* 1982; **36**: 269–73.
116. 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**: 67–9.
117. 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**: 355–61.
118. Jabine TB. Reporting chronic conditions in the National Health Interview Survey. A review of findings from evaluation studies and methodological test. *Vital Health Stat 2* 1987 Aug(105): 1–45.
119. Madow WG. Interview data on chronic conditions compared with information derived from medical records. *Vital Health Stat 2* 1967 May(23): 1–84.

120. Watson DL. Health interview responses compared with medical records. *Vital Health Stat 1* 1965 Jul(46): 1–74.
121. Wilcox AJ, Horney LF. Accuracy of spontaneous abortion recall. *Am J Epidemiol* 1984; **120**: 727–33.
122. 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**: 1–7.
123. Biering-Sorensen U. Reporting of hospitalization in the Health Interview Survey. *Vital Health Stat 1* 1965 (61): 1–71.
124. 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**: 875–9.
125. 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**: 186–9.
126. 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**: 206–17.
127. Udry JR, Gaughan M, Schwingl PJ, van den Berg BJ. A medical record linkage analysis of abortion underreporting. *Fam Plann Perspect* 1996; **28**: 228–31.
128. Lessler JT, Harris BSH. *Medicaid Data as a Source for Postmarketing Surveillance Information*. Research Triangle Park, NC: Research Triangle Institute; 1984.
129. Cannell CF, Marquis KH, Laurent A. A summary of studies of interviewing methodology. *Vital Health Stat 1* 1977(series 2): i–viii, 1–78.
130. 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**: 288–300.
131. Fourrier-Reglat A, Cuong HM, Lassalle R, Depont F, Robinson P, Droz-Perroteau C, *et al.* Concordance between prescriber- and patient-reported previous medical history and NSAID indication in the CADEUS cohort. *Pharmacoepidemiol Drug Saf* 2010; **19**: 474–81.
132. 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**: 1530–6.
133. 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**: 559–64.
134. Bonito AJ, Farrelly MC, Han J, Lohr KN, Lubalin JS, Lux L. *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, MD: Agency for Health Care Policy and Research, 1997.
135. Patient Protection and Affordable Care Act (PPACA). Sect. 111–148 (2010 March 23).
136. Stergachis AS. Record linkage studies for postmarketing drug surveillance: data quality and validity considerations. *Drug Intell Clin Pharm* 1988; **22**: 157–61.
137. Rawson NS, Harding SR, Malcolm E, Lueck L. Hospitalizations for aplastic anemia and agranulocytosis in Saskatchewan: incidence and associations with antecedent prescription drug use. *J Clin Epidemiol* 1998; **51**: 1343–55.
138. Schering Laboratories. *The Phantom Patient and Community Pharmacy Practice*, Schering Report SVII. Kenilworth, NJ, 1996.
139. Craghead RM, Wartski DM. An evaluative study of unclaimed prescriptions. *Hosp Pharm* 1991; **26**: 616–7; 32.
140. Farmer KC, Gumbhir AK. Unclaimed prescriptions: an overlooked opportunity. *Am Pharm* 1992; **NS32**: 55–9.
141. Hamilton WR, Hopkins UK. Survey on unclaimed prescriptions in a community pharmacy. *J Am Pharm Assoc* (Wash) 1997; **NS37**: 341–5.
142. Kirking MH, Kirking DM. Evaluation of unclaimed prescriptions in an ambulatory care pharmacy. *Hosp Pharm* 1993; **28**: 90–1, 4, 102.
143. Kinnaird D, Cox T, Wilson JP. Unclaimed prescriptions in a clinic with computerized prescriber order entry. *Am J Health Syst Pharm* 2003; **60**: 1468–70.
144. Bright RA, Avorn J, Everitt DE. Medicaid data as a resource for epidemiologic studies: strengths and limitations. *J Clin Epidemiol* 1989; **42**: 937–45.
145. Strom BL, Carson JL, Schinnar R, Snyder ES, Shaw M, Waiter SL. No causal relationship between transdermal scopolamine and seizures: methodologic lessons for pharmacoepidemiology. *Clin Pharmacol Ther* 1991; **50**: 107–13.
146. 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**: 1858–65.

147. Curtis JR, Cheng H, Delzell E, Fram D, Kilgore M, Saag K, *et al.* Adaptation of Bayesian data mining algorithms to longitudinal claims data: coxib safety as an example. *Med Care* 2008; **46**: 969–75.
148. Brophy JM, Levesque LE, Zhang B. The coronary risk of cyclo-oxygenase-2 inhibitors in patients with a previous myocardial infarction. *Heart* 2007; **93**: 189–94.
149. Varas-Lorenzo C, Castellsague J, Stang MR, Perez-Gutthann S, Aguado 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**: 1016–25.
150. Wahl PM, Rodgers K, Schneeweiss S, Gage BF, Butler J, Wilmer C, *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**: 596–603.
151. 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**: 1277–82.
152. Newton KM, Wagner EH, Ramsey SD, McCulloch D, Evans R, Sandhu N, *et al.* The use of automated data to identify complications and comorbidities of diabetes: a validation study. *J Clin Epidemiol* 1999; **52**: 199–207.
153. 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.
154. 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**: 299–306.
155. 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**: 1397–407.
156. Raiford DS, Perez Gutthann S, Garcia Rodriguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. *Epidemiology* 1996; **7**: 101–4.
157. van Staa TP, Smeeth L, Persson I, Parkinson J, Leufkens HG. Evaluating drug toxicity signals: is a hierarchical classification of evidence useful or a hindrance? *Pharmacoepidemiol Drug Saf* 2008; **17**: 475–84.
158. 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**: e1000194.
159. Andrade SE, Gurwitz JH, Chan KA, Donahue JG, Beck A, Boles M, *et al.* Validation of diagnoses of peptic ulcers and bleeding from administrative databases: a multi-health maintenance organization study. *J Clin Epidemiol* 2002; **55**: 310–3.
160. Donahue JG, Chan KA, Andrade SE, Beck A, Boles M, Buist DS, *et al.* Gastric and duodenal safety of daily alendronate. *Arch Intern Med* 2002; **162**: 936–42.
161. Dornan S, Murray FE, White G, McGilchrist MM, Evans JM, McDevitt DG, *et al.* An audit of the accuracy of upper gastrointestinal diagnoses in Scottish Morbidity Record 1 data in Tayside. *Health Bull* (Edinb) 1995; **53**: 274–9.
162. Evans JM, McMahon AD, McGilchrist MM, White G, Murray FE, McDevitt DG, *et al.* Topical nonsteroidal anti-inflammatory drugs and admission to hospital for upper gastrointestinal bleeding and perforation: a record linkage case-control study. *BMJ* 1995; **311**: 22–6.
163. Hammad TA, McAdams MA, Feight A, Iyasu S, Dal Pan GJ. 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**: 1197–201.
164. Garcia Rodriguez LA, Barreales Tolosa L. Risk of upper gastrointestinal complications among users of traditional NSAIDs and COXIBs in the general population. *Gastroenterology* 2007; **132**: 498–506.
165. 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, OH: Harvey Whitney, 1991, pp. 222–34.
166. 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**: 907–17.
167. Strom BL, Carson JL, Morse ML, LeRoy AA. The computerized on-line Medicaid pharmaceutical analysis and surveillance system: a new resource for

postmarketing drug surveillance. *Clin Pharmacol Ther* 1985; **38**: 359–64.

168. Ray WA, Griffin MR. Use of Medicaid data for pharmacoepidemiology. *Am J Epidemiol* 1989; **129**: 837–49.
169. Iezzoni LI, Burnside S, Sickles L, Moskowitz MA, Sawitz E, Levine PA. Coding of acute myocardial infarction. Clinical and policy implications. *Ann Intern Med* 1988; **109**: 745–51.
170. Pladevall M, Goff DC, Nichaman MZ, Chan F, Ramsey D, Ortiz C, *et al.* An assessment of the validity of ICD Code 410 to identify hospital admissions for myocardial infarction: The Corpus Christi Heart Project. *Int J Epidemiol* 1996; **25**: 948–52.
171. An act making supplemental appropriations for job preservation and creation, infrastructure investment, energy efficiency and science, assistance to the unemployed, and State and local fiscal stabilization, for the fiscal year ending September 30, 2009, and for other purposes. S. Rept. 111-3 ARRA (2009 February 17).

CHAPTER 42
Studies of Medication Adherence

Trisha Acri[1] *and Robert Gross*[2]

[1] Department of Family and Community Medicine, Temple University School of Medicine, Philadelphia, PA, USA

[2] Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

In this chapter, we will describe the importance of the measurement of adherence in pharmacoepidemiologic research, the methods by which medication adherence can be measured, methodologic issues that arise once adherence has been measured, and future directions for the field. While we use many different drug–disease examples in this chapter, we focus heavily on examples from antiretroviral therapy for HIV disease because it has been at the forefront of adherence research.

Data show that as many as half of patients do not take all of their prescribed medication[1] resulting in more than $100 billion of avoidable hospitalizations.[1] Improvements in adherence to antihypertensives could help prevent more than 89 000 premature deaths in the United States alone each year.[1] However, without accurate measurement of adherence incorporated into research and clinical practice, the problem will remain underappreciated and poorly addressed.

The various definitions used to describe the behavior of interest can be confusing. *Compliance* has been defined as "the extent to which the patient's dosing history conforms to the prescribed regimen."[2] The idea of a patient passively "conforming" to the prescriber's will is often viewed as imposing a judgmental framework on the problem; therefore, the term *adherence* has come to supplant the term compliance. *Adherence* more strongly conveys the idea of a treatment alliance or relationship where the patient implements the recommendations of the provider.[3] But, adherence to a regimen encompasses several steps that result in the patient actually consuming the medication. *Acceptance* denotes initial engagement with the prescribed medication.[4] *Persistence* refers to how long the patient continues to follow the regimen.[4] *Execution* represents how well the patient follows the prescribed regimen during the time s/he is engaged with treatment.[4] Each step is required in the process; failure to perform any of these results in non-adherence. These terminologies highlight the fact that the patient who never fills a prescription, perhaps because of the cost of the medication, differs from the patient who misses an occasional dose because of forgetfulness, and differs from the patient who begins the therapy and executes it well at first but does not continue to take it over time. Defining these three patients as non-adherent based on the average percentage of medication consumed over a certain time period ignores the likelihood that each of these patients might have somewhat different treatment outcomes and would require different interventions to improve upon their adherence.[5] However, a single percent of doses taken over the entire time period is often used as the sole measurement of adherence in published research.

The actual behaviors involved in taking a prescribed medication as directed become much more

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

complicated when each step of the process is considered. This is just one of many limitations encountered in the study of adherence and why measuring adherence and attempts to enhance adherence are so difficult. Regardless, practical approaches to measuring and analyzing adherence have been developed. We will discuss the utility of the varied approaches and the remaining challenges.

Clinical problems to be addressed by pharmacoepidemiologic research

Adherence research confronts the truism attributed to US Surgeon General C. Everett Coop, MD that "drugs don't work in patients who don't take them."[3] As such, measurement of adherence is essential in order to address several issues in the interpretation of studies of beneficial and adverse effects of medications. In randomized controlled trials, adherence to treatment can be an important factor affecting the outcome of the trial and the resultant estimates of efficacy and safety of the tested medication. Poor adherence to the drug being tested can lead to underestimates of the efficacy of the drug being tested.[6] Further, adherence information allows for a more accurate assessment of the incidence of toxicity from a drug because those who do not take the drug cannot have toxicity from it. Since no one can be expected to have perfect adherence all the time, including clinical trial participants, the measurement of adherence can help to inform whether a drug fails to exert an effect because it did not work or because it was not taken.

Once a medication is marketed, the pattern of prescribing by clinicians is only part of the picture of how drugs are used by populations of patients. Patients who volunteer for clinical trials are probably more motivated to take the medication than those given prescriptions as part of usual care. Therefore, measuring adherence in observational studies of drug effectiveness and toxicity may be even more important than in clinical trials. Furthermore, the advantage of assessing adherence in observational studies is that it provides a more "real-world" estimate of adherence in clinical populations than in the artificial setting of a clinical trial. Moreover, since it is a major determinant of treatment outcome for chronic diseases managed with efficacious medications, adherence itself can be the focus of pharmacoepidemiologic research.

Non-adherence can be volitional or unintentional. Observational studies of barriers to adherence over the past several decades have identified many potential barriers to adherence. These can be categorized as patient-level factors, system-level factors, and medication-specific factors. Common patient barriers consist of forgetting to take the medication,[7] missing doses to avoid side effects,[7,8] and psychosocial factors such as depression and lack of social supports.[8] Common system barriers include logistical difficulty in obtaining the medication from the dispenser and, in some settings, sporadic drug unavailability ("stock outs").[9] Common medication-specific barriers are dosing frequency[8] and adverse effects.[7,8] Patients in observational settings miss doses for the same reasons as participants in clinical trials, but may also be affected more by lack of trust[10] or lack of motivation[11] because they differ from clinical trials volunteers. Further, patients may decide on a dose-by-dose basis whether to take medicine as prescribed for a variety of different reasons. For example, patients may take doses intermittently to avoid side effects at particular times[7,8] (e.g., avoidance of diarrhea when needing to take public transportation). In addition, adherence may wane over time and post-marketing observational studies may demonstrate pill fatigue when patients are followed for longer periods of time than is typically done in clinical trials.[12] Thus, adherence studies in observational settings can provide unique data not available from trials.

While missing doses is the more common adherence problem, taking extra doses can also be an issue in some settings. Drugs with a narrow therapeutic window are also of concern here where extra doses may result in toxicity. An example is warfarin for anticoagulation.[13] Of course, patients may take extra doses of narcotics prescribed for the treatment of pain because of inadequate pain relief or for potential abuse.

Adherence is important not only for the patient's clinical outcome, but can also impact public health, particularly regarding infectious diseases. In tuberculosis and HIV, non-adherence actually modifies the disease itself by selecting for organisms that are resistant to the treatment. Since these diseases are transmissible, and transmitted resistance has been confirmed,[14–16] the measurement of non-adherence and interventions to improve it for the individual take on greater importance.

Measurement of adherence can also be useful for determining the threshold of how much medication must be taken to obtain the desired clinical outcome. For example, oral contraceptive guidelines describing the use of double doses and need for backup methods of contraception when doses are missed[17] were informed by studies assessing the effects of treatment interruptions, which are essentially periods of non-adherence.[18] These dosing thresholds likely differ by drug and disease. In hypertension, taking at least 80% of prescribed medication is an acceptable standard for blood pressure control.[19] Yet in HIV, this standard is often insufficient for treatment success. In a study of patients newly starting protease inhibitor therapy for treatment of HIV, those who took 80–95% of doses were more likely than those with lower adherence rates to achieve complete suppression of viral replication.[20] Unfortunately, such detailed information is not available for most drugs and diseases. Therefore, the default goal regarding adherence should be to encourage the patient to take as many prescribed doses as possible.

Methodologic problems to be solved by pharmacoepidemiologic research

Challenges in the measurement of adherence

The gold standard for measuring adherence to pharmacotherapy is directly observed therapy. However, this approach is only practical in limited settings, such as the administration of a novel agent in a controlled environment. Provider predictions of adherence have been shown to be no better than chance[21] and therefore should not be used. While many approaches to measuring adherence exist, as we will discuss later, whatever the approach, the discovery of non-adherence can be embarrassing for the patient. Non-adherence may be stigmatizing because it implies lack of respect for the advice given by a provider or lack of caring for one's own welfare. Thus, knowledge that one's adherence is being monitored, unless done unobtrusively, risks influencing the behavior it is meant to measure (i.e., a Hawthorne effect). In addition, tracking of a daily activity can be seen as burdensome whether or not individuals are aware of their own non-adherence. Therefore, the measurement of adherence requires creative approaches to assess a daily activity performed at different times per day for different individuals.

Challenges in the analysis of adherence data

Once adherence is measured, the best approach to analyzing the data becomes central. In clinical trials, adjusting results for adherence is complicated by the fact that adherence itself is related to better health outcomes, irrespective of receiving active drug or placebo. In the Beta-Blocker Heart Attack Trial (BHAT), a randomized double-blind placebo-controlled trial of propranolol after myocardial infarction, the odds ratio of mortality in poor adherers in the active arm (who took less than 75% of propranolol) compared with good adherers was 2.8 after adjustment for potential confounding factors. Notably, the adjusted odds ratio of mortality in the group with poor adherence to placebo was, similarly, 2.7. Presumably, adherence to the medication, whether propranolol or placebo, was strongly associated with other factors (e.g., lifestyle) related to mortality. This is particularly relevant in the setting where other uncontrolled factors, such as diet and exercise, play a role in determining outcome.[22] How to control for this effect is an important issue in the analysis of such studies.

Other analytic challenges include the duration and timing of adherence measurement. Adherence behavior varies over time. Therefore, for chronic treatments, an individual may be adherent for part of the observation period and non-adherent for

another. For example, when assessing the outcomes of the treatment for HIV, individuals are prescribed lifelong regimens. In any one year, adherence over the initial 12 weeks of observation may not be the same as adherence over the final 12 weeks. Therefore, simply summing adherence over the entire 48-week interval will provide an average amount of adherence for the treatment course. Yet, short periods of non-adherence, as seen in treatment interruption studies,[23] can have a major impact on outcome. Further, whatever the interval chosen, the summation of the adherence data during that interval can be accomplished in many different ways. The simplest is the percent of doses taken; however, this might not be the most clinically relevant metric. Depending on the pharmacokinetics and pharmacodynamics, duration of gaps and variability in adherence over time may be more important than the simple proportion of prescribed doses taken. Defining the duration of the adherence interval, determining which intervals are likely to be related to the treatment outcome, and defining the metrics of interest are key issues in the analysis of this phenomenon.

Additionally, many diseases are treated with combination therapy. While these drugs are often studied in combination to determine their effect (e.g., antihypertensive therapies, antituberculous therapies), it is challenging to determine how to weight differential adherence among the drugs. Many of these issues can be addressed with currently available solutions, although methodologic challenges remain to be solved.

Currently available solutions

Overview

There are many different methods for measuring adherence to medication, 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 measurements provide more granular data with respect to timing of dose-taking. For example, in clinical trials, because of the frequent patient contact and the prospective nature of the study, many of the techniques we will discuss may be used. In other settings, such as retrospective studies using databases, the options become more limited. One strategy we will not discuss is provider estimation. Providers are poor predictors of future adherence and poor estimators of previous adherence. For example, in a study assessing adherence to protease inhibitors in patients newly initiating antiretroviral therapy, provider predictions of future adherence over the next 4 months and provider estimates of adherence during the 4-month study were compared to microelectronic drug monitors. There was no correlation between actual and predicted adherence.[21]

Once adherence is measured, there are several issues that must be addressed. As discussed above, adherence is a time-varying phenomenon. Therefore, measurements must not only be made serially, but also need to address the question of how long a period of time should be considered an adherence "interval." Furthermore, because the effects of drugs can occur well after they are stopped and the offset times of drugs differ depending on pharmacokinetics and pharmacodynamics, the timing of the measurement must also be carefully considered.

We will describe each of the strategies and their strengths and weaknesses and then discuss general considerations for the timing of assessing adherence and the duration of an adherence interval.

Specific techniques for measuring adherence

Self-reports

Self-reports of medication adherence are used most commonly in clinical practice because they are simple, relatively inexpensive, and feasible. They can be obtained over the telephone, in person, or with paper or electronic surveys. Several different methods for assessing self-reported adherence have been found to be valid measures of adherence. The Adult AIDS Clinical Trials Group instrument queries subjects about the number of missed doses of each medication over each of the last 4 days, missed doses on the weekends, the last missed dose, and

adherence to dietary instructions. The form is self- or interviewer-administered and takes participants an average of 10 minutes to complete.[7] The instrument has been modified to include reports of adherence over the last 3 days[24] and to incorporate the last time a dose was missed.[25] A simpler but still comprehensive measure validated in a large cohort of patients with HIV that can be used for other medications is the Simplified Medication Adherence Questionnaire, which asks patients specific questions about forgetfulness or carelessness about taking medications and then asks them to report missed doses over the previous 24 hours, the past week, the last weekend, and the last 3 months.[26] 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.[24] Still others ask participants to estimate numerically how much medication they have taken over a specified time period. These methods, which can be self-administered in high-literacy patients or can be conducted by an interviewer, are all limited by a patient's ability to recall missed doses and biased by social desirability (reporting conformity with physician instructions to avoid embarrassment). Social desirability can be mitigated by using permissive statements like, "We know that it is sometimes difficult to take all your medications on time as directed."[7]

Simple interview techniques are potentially limited by multiple factors including language barriers, literacy, time burden to the provider, social desirability, and difficulty with communication of complicated regimens and medication names. Use of audio computer-assisted self-administered interview (ACASI) can reduce these barriers. The computer-aided strategy can utilize an audio track to read instructions and questions and can include high-resolution photographs of the medicines rather than medication names to assist participants with lower literacy. The adherence questions can be administered at a kiosk or computer in a physician's waiting room to minimize burden to the provider in the office and to decrease social undesirability of admitting poor adherence to clinical personnel. Empirical data suggest that computer-aided self-reports are less likely to overestimate adherence.[27] However, the issue of faulty recall is not resolved by this approach.

Lu *et al.* conducted a comparison study of 3-day, 7-day, and 1-month self-reports of adherence and showed that for all three time periods, estimates of adherence were significantly higher than those obtained using electronic drug monitors (EDM), suggesting that self-reports of adherence generally overestimate how much medication has been taken.[28] In this study, however, the 1-month estimates of adherence best approximate adherence measurements obtained using EDM.[28] A review of the literature by Simoni and colleagues on use of self-reports for measuring adherence to antiretroviral therapy showed that the most commonly used measure was a single item asking participants to specify how much medication had been missed over a specified time period varying between 2 and 365 days. The review suggested that querying patients about their medication taking behavior over longer periods of time may be more strongly associated with clinical response than shorter periods.[29] These studies together suggest that querying participants about time periods of adherence of a month or more when using self-reported measures will better predict both EDM adherence and viral load outcomes.[28,29]

Refill data

Pharmacy refill data was pioneered by Steiner *et al.* in the late 1980s[30] and has been widely used in various chronic diseases to date. The high quality of pharmacy refill data is predicated on a pair of positive and negative incentives for its accuracy. If refills are dispensed, but not recorded, the dispenser does not get reimbursed for the medication. If refills are recorded, but not dispensed, the dispenser is guilty of fraud. The data quality may be less assured in settings where such tracking is less crucial for reimbursement. In contrast to self-reports, the pharmacy refill measure is less susceptible to deception, not biased by poor recall, can be obtained from computerized records, and can be assessed retrospectively.[31]

The pharmacy refill measure of adherence has been described and used with several different

methods and definitions. In all approaches, the dates on which refills of medication are obtained from the pharmacy and the duration of the supply dispensed are used to estimate adherence. Whichever approach is considered, this strategy assumes that an individual would not obtain a new refill in the pharmacy until all doses of the prior bottle of medication have been consumed. Further, it requires that the medication be prescribed for chronic use, that is, for a condition requiring more than a single prescription.

Although the pharmacy refill measure of adherence estimates the amount of medication an individual has in his/her possession during a given time period, it does not measure or monitor actual pill-taking behavior, either on average or day-to-day. As such, it cannot be used when the timing of missed doses during the interval is pivotal. However, the technique is a valid measure of adherence for chronic medications where measuring exposure between refills is clinically relevant. For example, a time-to-refill measure of adherence has been associated with changes in viral load in HIV[31] and changes in blood pressure in hypertension.[30] Furthermore, the measure has been shown to provide additional information over and above self-reported adherence data. In a study of antiretroviral therapy, individuals self-reporting 100% adherence actually varied in their treatment response based on their adherence as measured using the refill technique. As expected, those with higher adherence on the refill measure had higher rates of treatment response, despite claims of perfect self-reported adherence in both groups.[31]

There are limitations to the use of this measure of adherence in that it gives only an estimate of the maximum amount of medication the person can have in his/her possession. The measure is most feasible when prescription refills are obtained from a single pharmacy or the information is obtained centrally in a data repository such as may be used with managed care insurance. The accuracy of pharmacy refill data for measuring adherence can also be limited by the possibility that patients might obtain medication from other sources such as friends or family, or during hospitalization. It is also possible that prescribed dosages are changed during the time of monitoring. An attempt to define these details of medication-taking through self-reports can better inform measurements of adherence.[30] Notably, it is not useful when refills are automatically mailed to patients without their needing to contact the dispenser.

In the most commonly applied approach, the proportion of doses taken over the interval is calculated using the following formula: days supply/(last day of refill minus first day of refill).[31] Other approaches include calculating the number of refills obtained divided by the number of refills expected over the time frame, or calculating the amount of time between refills as a gap in adherence. Although easier to calculate, the simpler measures are more limited in their utility. For example, the number of 1-month refills divided by the number of months of follow-up over time[32] is difficult to use over a shorter time period (e.g., a few months). The problem can be seen when assessing a 3-month refill interval. Since three 30-day refills are needed in this situation, adherence can only be categorized as 100% (3/3), 66% (2/3), 33% (1/3), or 0% (0/3). In this scenario, the patient who returns for the last refill 1 day late would be assigned 66% adherence. The time-to-refill approach addresses this issue. Over a 3-month time period during which three refills of 30 days' supply each would be expected, the patient's adherence would be 90/91 or 99%, a much better reflection of actual medication consumption than 66%. This measure gives greater sensitivity, and can be particularly useful when adherence differences between 66% and 100% are crucial, as in HIV disease outcomes.[20] Of course, this problem becomes less important the longer the interval that is being assessed. However, the shorter the interval assessed, the more rapidly can nonadherence be identified and acted upon.

Refill-based measures must always address the issue of non-refill. This measurement technique requires that the days supply be determined and then divided by the time period of interest. However, if the patient stops taking the medication, the recorded stop date is either inaccurate or, more likely, unavailable. In this scenario, it is necessary

to artificially assign a refill date for the last refill to close the time interval. Potential approaches include setting it at the end of follow-up or at a fixed point after the last refill (e.g., 2 months late for a 1-month supply). Whichever method is chosen, sensitivity analyses in which this artificial stop date is assigned different times (often the extremes of possibility) make the results more robust.

Pill counts

Adherence can also be measured by pill counts. Pill counts are similar to pharmacy refill data in that percent adherence is calculated by dividing the days supply consumed by the number of days observed. Like refill data, pill counts cannot determine if the medication was actually taken or the pattern of medication taking. However, they do provide direct evidence that the medication was not taken when pills are left over. Pill counts are more susceptible to patient 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 alternatives to mitigate this type of misclassification. Keeping the individual unaware of when the count will occur and having the count occur when pills are expected to remain in the bottle decrease the likelihood of pill dumping. Unannounced pill counts can occur during office visits or home visits. This approach has been validated in studies of homeless and marginally-housed adults with HIV in San Francisco.[33,34]

While unannounced pill counts decrease the risk of pill dumping, they can often be logistically infeasible for staff, especially if needed to be done in the home. Modifying the approach to performing pill counts by unannounced telephone call is a potentially more feasible alternative. Subjects are told these telephone calls will occur periodically and are trained at the outset in-person to conduct their own pill counts. At the time of the call, they bring all medication bottles to the telephone and the content of each bottle is reviewed individually. Data collected include the date of fill, the quantity dispensed, the number of pills per dose, and the number of pills left in the bottle. The counts are adjusted for doses taken that day and for any additional pills left over from the last pill count. Of course, this approach is also susceptible to intentional deception. Yet, it has been found to be associated with treatment response in an HIV cohort and estimated to take approximately 18 minutes per participant.[35]

The time and annoyance for both the staff who do the counting and the participants who need to bring their pill bottles with them are potential disadvantages of pill counting and an additional source of error. Irrespective of the approach, pill counting is frequently viewed as burdensome by study personnel. Reinforcing the importance of accuracy with the staff is vital for this measure to be valid.

Medication diaries

Although the measures described above yield a global amount of drug taken over a specified time period, the measures give no detail on the timing of the medication-taking and missed doses. Depending on the pharmacokinetics and pharmacodynamics, missing doses several days during a month may have different consequences depending on whether the doses were missed consecutively or if they were missed at separate times spaced evenly throughout the month.[4] These data may in fact be vital to the adherence classification. Medication diaries can provide this information. With this technique, 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. This information can be critical to interpreting results of studies of drugs that are difficult to track using other strategies, such as insulin.[36]

Medication diaries are susceptible to both over-reporting and under-reporting of adherence. Social desirability results in patients listing doses even though they were not taken. This potential for deception is lessened somewhat by the burden of needing to create a detailed falsified record. In fact, the burden of tracking each dose actually increases the risk of under-reporting since medications may be taken without access to the diary or the process of writing down the dose may be forgotten or ignored.

Electronic drug monitors

Electronic drug monitors (EDMs) feature the same advantages as the medication diaries, but are less susceptible to deception, forgetting, or ignoring the need to write down the dose data. While there are several different hardware options, electronic drug monitors employ electronic date/time stamp technology that is triggered by the process of opening the container or puncturing a blister pack to obtain the dose. The data are downloaded to a computer via hardwired or wireless linkage for analysis. As stated above, EDMs are less susceptible to deception. While not impossible, it is thought to be the rare subject who will game the system by opening and closing the monitor to record medication-taking events over long periods of time without actually taking the medication. They are also less susceptible to under-reporting than medication diaries because they often do not require the subject to do anything other than take the prescribed medication. Additionally, they can be used as medication diaries even when the medication is not kept in the container. In a study of warfarin adherence, individuals using medication organizer boxes were given an EDM on an empty pill container and were asked to open the empty bottle whenever they took their warfarin from the organizer box. The association between adherence and outcome was nearly as strong for this subset of individuals as for those who kept the warfarin in the monitored bottle itself.[13]

The packaging of EDMs can be burdensome. They often preclude the use of pill boxes and require that the medication remain in the package until taken. Therefore, they are susceptible to underestimating adherence (e.g., a 1-week supply of doses taken from the container at one time will appear as one dose taken). Another limitation of this approach is the cost of purchasing the monitors and the hardware and software to analyze the data, and this limits their use in clinical practice.

Drug concentrations

Identification of the presence of a drug in plasma or other tissues provides evidence of drug ingestion. However, the use of drug concentrations to measure adherence is limited by variability between patients in drug processing (i.e., absorption, distribution, metabolism, and clearance; see Chapter 2). The more variable those steps are between patients, the weaker the relation between drug concentration and adherence. Further, if the drug has a short half-life in the compartment (e.g., plasma), then the measurement only captures short-term use. Thus, this approach is limited by the issue of interval censoring between measurements. The more frequently these are measured, the fuller the picture of adherence behavior that can be obtained. But, then cost and inconvenience to the patient may come into play.

Measurement of drug concentrations in hair assessed by liquid chromatography and confirmed by tandem mass spectrometry can be a useful indicator of long-term exposure to medication. For example, antiretroviral drug levels in hair give an average of the exposure to drug over the past weeks to months and operate better than serum drug levels to predict HIV viral response.[37,38] The duration of exposure to a drug measured by hair drug concentrations and serum drug levels is analogous to the relationship between hemoglobin A1C and a single serum glucose measurement.[39] Hair concentrations have been used frequently for forensic purposes to determine exposure to nonprescription drug abuse. There is interest in using hair analysis with antiseizure medications and psychotropic drugs as well. In HIV therapy, the technique has been demonstrated in indinavir,[37] lopinavir/ ritonavir,[38] and atazanavir alone and with ritonavir.[38] The method is being developed for use with efavirenz.[40] Gandhi and colleagues demonstrated that drug concentrations in hair predicted viral outcomes better than self-reported adherence.[38]

Unfortunately, many of these assays are unavailable commercially. Furthermore, for other drugs, the serum drug level is not the relevant measure because the site of action is elsewhere. For example, nucleoside analog reverse transcriptase inhibitors act intracellularly; therefore, concentrations in the cells would be the relevant measures but these are very challenging to assay.[41] Turnaround time is another issue. Unless these assays are done in real time or nearly so, they are not useful clini-

cally. If samples are batched over weeks to months, the non-adherence would be detected long after the fact. The issue of multiple drugs in a regimen is also important here as it is for other measures. Assays for each drug may not be available, and the classification of an individual as adherent or not is difficult if some of the drug concentrations are in an acceptable range while others are not.

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.[42] Fluorescence spectrophotometry is used to assess the concentration. Of course, this strategy is only relevant in settings where control over the formulation is in the hands of the researchers (i.e., clinical trials).

Adherence to non-pill formulations

Medication diaries and self-reports can be used to monitor adherence to non-pill formulations of medication therapy, but are subject to the same biases described above for pill formulations. Particular circumstances raise several unique challenges to measuring non-pill formulation adherence. For example, measuring adherence to injectable pegylated interferon for the treatment of hepatitis C is feasible using pharmacy refill dates and the number of interferon syringes dispensed with each refill,[43] because the days supply is fixed and/or syringes are pre-filled. However, when an injectable medication such as insulin is administered based on a sliding scale, with doses adjusted as needed, measuring adherence using refill data may be invalid.[44]

Topical treatments pose a particular challenge. For transdermal formulations in patches (e.g., nicotine, testosterone), because the supply is typically fixed, refill adherence is a viable option. 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 or other characteristics, self-reports and medication diaries may be the only viable options at present.[45]

Adherence to intravaginal gels can be monitored by counting the number of empty tubes and used applicators returned at each study visit,[46] but this measure is subject to self-report errors due to intentional falsification or mixture of used and unused applicators in the same bag.[47] Assays using trypan blue have been piloted to improve adherence assessments in trials testing the use of intravaginal microbiocides to prevent sexual transmission of HIV. The trypan blue, when applied to returned vaginal applicators after reported use for medication administration, will detect exposure to cervical mucous and differentiate between used and unused applicators. The assaying dye can be administered to the returned applicators in a plastic bag or a shallow pan or can be aerosolized and sprayed over larger batches of applicators and will preferentially stain applicators that have been inserted intravaginally.[48] Because of the potential carcinogenic and toxic effects of trypan blue when applied as an aerosol to large batches of applicators during clinical trials, the use of blue food dye, which also preferentially stains applicators with cervical mucous present, was tested for a similar purpose with high accuracy.[47]

EDMs have been used for metered dose inhalers[49] and ophthalmologic solutions.[50] The monitors do increase the bulk of the packaging. However, unlike pills, these formulations cannot be taken out of the package. Thus, the patient burden of needing to keep pill formulations in the monitored package is not relevant for these formulations.

Analysis issues

Use of adherence data in the interpretation of clinical trials

While clinical trial participants may be more motivated to adhere than their counterparts in the clinical setting, non-adherence occurs in the settings of all self-administered therapy. Missed doses will typically make the active drug less effective and diminish the observed difference when compared to placebo. In order to compensate for this effect, trials may inflate sample sizes to account for this variability in drug exposure.[6] Clinical trials may also incorporate run-in periods to try to minimize poor adherence (see Chapter 36).

In analyzing trials in which non-adherence occurs, the standard approach remains intention-to-treat. This approach limits the introduction of bias and makes the results more generalizable to clinical practice.[51] Secondary analyses can be done, limiting inclusion exclusively to the adherent subset. Unfortunately, in the setting in which lifestyle changes serve as co-interventions with medication (e.g., studies of treatment of congestive heart failure), such secondary analyses only tease out the effect of the drug over and above lifestyle change and are not true measures of drug efficacy. Of course, as with all such secondary analyses, the benefits of randomization are negated and results can sometimes be difficult to interpret.

This difficulty was seen in the results of the Coronary Drug Project Research Group in which men who had previously had a myocardial infarction were randomized to receive either clofibrate or placebo. A subgroup analysis of high adherers, defined as those subjects who took at least 80% of the prescribed dose of either treatment or placebo, demonstrated no difference in overall mortality (15.0% for clofibrate and 15.1% for placebo). In the subjects with less than 80% adherence, mortality was lower in the clofibrate group (24.6%) compared to the placebo group (28.2%).[52] Therefore, it appears that some exposure to clofibrate could be better than none when overall adherence is poor. But, when adherence to medication treatment is higher, adherence to other co-interventions such as diet and lifestyle changes might outweigh the drug effects. Measurement of adherence for this type of analysis clearly enriches the understanding of the effects of the drug. Yet, the overall effectiveness of the medication is difficult to ascertain. The presence of an effect modifier, in this case adherence, typically makes interpretation of the results more complex.[51]

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 compound retains any potential for further use. For example, in the Lipid Research Clinics Coronary Primary Prevention Trial, rates of coronary heart disease events were compared in participants with hypercholesterolemia who were randomized to receive either cholestyramine to decrease lipid levels or a placebo.[53] Adherence to treatment 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 in the treatment arm. As a result of low adherence, the lipid-lowering response and the decrease in cardiac events in the treatment group were attenuated, and the difference in event rates in a comparison between the two groups was lower than it could have been if adherence were higher. 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.[51]

Time-varying nature of adherence and duration of adherence intervals

Adherence is a dynamic phenomenon. Therefore, categorizing an individual as adherent or not requires that the time interval of interest be specified. For chronic medications, using a single adherence metric to categorize an individual's behavior over long time periods may not be relevant. For example, an individual on antiretroviral therapy who interrupts treatment for as little as 2 weeks is likely to experience virologic failure.[23] However, adherence metrics that summarize medication taking over a year would yield very small deviations from perfect adherence (50 weeks adherent/52 weeks of observation = 96%). In contrast, if 1-month intervals were chosen, the individual would have 11 months of perfect adherence and 1 month of 50% adherence. This low adherence month would explain the treatment failure more clearly than does 96% adherence.

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 off-set of action, short intervals are likely to be more clinically relevant than when the drugs have long half-lives and longer off-sets of action. For adherence measures that can accurately assess adherence over short periods of time such as electronic data monitors, shorter intervals, when desired, can be calculated. In contrast, when measures such as pharmacy refills are used, intervals can only be as short as the expected time between fills (e.g., 30 days).

The relation between adherence and outcome has been well described in antiretroviral therapy and oral contraceptives. Using refill data, intervals of adherence as long as 1 year[32] and as short as 30 days[54] have been associated with viral load outcomes with antiretroviral therapy. In a direct comparison, a 90-day measure was found to be more strongly associated with viral load than a 30-day measure.[54] For oral contraceptives, two consecutive days of non-adherence resulted in an unacceptably high rate of treatment failure.[18]

Unfortunately, for the vast majority of medications, the duration of a relevant adherence interval is unknown. Much research is needed to optimize the assessment of adherence in chronic diseases such as diabetes, hypertension, and hypercholesterolemia. While the choice of interval length depends on the goals of the research, in general, monitoring adherence over shorter intervals is desirable. The shorter the adherence interval monitored the more readily barriers to adherence can be assessed and the more rapidly interventions can be implemented. But, this advantage comes at the expense of decreased accuracy regarding true adherence behavior. By way of illustration, in the extreme case, the time interval could be 1 day in which one missed dose would categorize an individual as non-adherent. Clearly, such a short interval is prone to misclassification of adherence status. But without information on the relation between adherence to the medication of interest and treatment outcome, investigators must choose without direct guidance. In these cases, choices for an adherence interval should be made based on pharmacokinetics and pharmacodynamics data (see Chapter 2).

Adherence metrics

The simplest approach to summarizing adherence is the percent of doses taken. However, this metric does not capture potentially relevant patterns of adherence. Other approaches are possible, particularly for electronic monitors and refill data. For electronic monitors, because the timing of each dose is available, percent of doses taken "on time," standard deviation of time between doses, duration of maximum time gap between doses, and many others can be calculated.[4,20,55] For refill data, metrics focus on either the percentage of available medication or the duration of gaps between refills.[30] Self-reports either focus on the proportion of doses the patients have taken or the time since the last dose was missed.[25]

Whichever metric is used, the choice as to whether to include adherence as a continuous or dichotomous variable depends on the question. Threshold levels must take into account both the likelihood of failure and the clinical consequences of treatment failure.[5] Few thresholds have been established based on evidence. Rather, 80% of doses taken is an often quoted magnitude to categorize good versus poor adherence, based upon expert opinion.[56] In other settings, higher magnitudes of adherence have been more closely associated with classifying individuals as having treatment success or failure.[43] Combination therapy potentially complicates this issue significantly. Since the amount of non-adherence to one drug can differ from another in the regimen, the categorization of an individual's adherence can be essentially infinite. Fortunately, there is some evidence to suggest that for medications taken simultaneously, adherence to one is highly collinear with adherence to the other.[57] Yet, differential non-adherence has been documented.[58]

Determining whether an individual is non-adherent or that the medication is no longer being prescribed can be difficult for all methods of measuring adherence. Information detailing the physician recommendation to stop taking the medication is needed to determine if such an individual is non-adherent. Further, even when these records are available and the provider documents the recommendation to stop the medication, the exact date can be difficult to determine.

The future

Although it was once thought that once-daily therapies would solve the problem of non-adherence, they have not. Nor have simple interventions worked. Furthermore, although risk factors for non-adherence have been useful in understanding the behavioral and structural factors underpinning non-adherence, no accurate and reproducible predictive model for non-adherence exists. Therefore, better methods for detecting non-adherence and better methods for addressing it will be welcome developments.

Regarding the measurement of adherence, novel strategies include the use of more advanced microelectronic technology, often linked with communication systems that both identify and report non-adherence. Mobile telephone technology has resulted in numerous developments for tracking adherence including short message system (SMS or "text messaging") and interactive voice messaging. Since telephones can also be used to speak to participants, these measurement tools are often used in the setting of telephone-based interventions. Refinements to currently available electronic monitors will likely include more convenient packaging that can both help with adherence (e.g., a reminder or organizer system) and serve as an electronic tracking system.[59] In addition, these systems have the potential to include two-way communication to both gather data and provide automated customized feedback.

Completely novel approaches are likely to emerge as well. Researchers have designed a prototype of a biocompatible digestible antenna and microchip. The antenna sends a signal to a device worn by the patient, and the device sends a signal to the doctor or research team monitoring adherence.[60] Another novel approach is the use of a marker in the pill coating that can be detected on the breath by sensor. Adherence data of the date and time the dose was reported to have been taken are captured electronically.[61]

Many challenges remain. Objective measurement of adherence to non-pill formulations in particular is difficult, especially for liquids and topical treatments. Regarding analyses, the optimal adherence metric for virtually every drug–disease dyad remains unknown. This is further complicated by the enormous number of possible combinations of partial adherence to drugs in combination regimens. Hopefully, with greater recognition of the importance of non-adherence, 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 efficacious medications can ultimately be maximally effective.

References

1. Cutler DM, Everett W. Thinking outside the pillbox—medication adherence as a priority for health care reform. *N Engl J Med* 2010; **362**: 1553–5.
2. Urquhart J. Defining the margins for errors in patient compliance. *Pharmacoepidemiol Drug Saf* 2000; **9**: 565–8.
3. Osterberg L, Blaschke T. Adherence to Medication. *N Engl J Med* 2005; **353**: 487–97.
4. Vrijens B, Vincze G, Kristanto P, Urquhart J, Burnier M. Adherence to prescribed antihypertensive drug treatments: longitudinal study of electronically compiled dosing histories. *BMJ* 2008; **336**: 1114–17.
5. Steiner JF, Earnest MA. The language of medication-taking. *Ann Intern Med* 2000; **132**: 926–30.
6. Ickovics JR, Meisler AW. Adherence in AIDS clinical trials: a framework for clinical research and clinical care. *J Clin Epidemiol* 1997; **50**: 385–91.
7. Chesney MA, Ickovics JR, Chambers DB, Gifford AL, Neidig J, Zwickl B, *et al.* Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: the AACTG adherence instruments. *AIDS Care* 2000; **12**: 255–66.
8. Ammassari A, Trotta MP, Murri R, Castelli F, Narciso P, Noto P, *et al.* Correlates and predictors of adherence to highly active antiretroviral therapy: overview of published literature. *J Acquir Immune Defic Syndr* 2002; **31**: S123–7.
9. Oyugi JH, Byakika-Tusiime J, Ragland K, Laeyendecker O, Mugerwa R, Kityo C, *et al.* Treatment interruptions predict resistance in HIV-positive individuals purchasing fixed-dose combination antiretroviral therapy in Kampala, Uganda. *AIDS* 2007; **21**: 965–71.
10. Altice FL, Mostashari F, Friedland GH. Trust and the acceptance of adherence to antiretroviral therapy. *J Acquir Immune Defic Syndr* 2001; **28**: 47–58.

11. Cooperman NA, Arnsten JH. Motivational interviewing for improving adherence to antiretroviral medications. *Curr HIV/AIDS Rep* 2005; **2**: 159–264.
12. 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**: 628–36.
13. Kimmel SE, Chen Z, Price M, Parker CS, Metlay JP, Christie JD, *et al.* The influence of patient adherence on anticoagulation control with warfarin. *Arch Intern Med* 2007; **167**: 229–35.
14. Yerly S, Kaiser L, Race E, Bru JP, Clavel F, Perrin L. Transmission of antiretroviral-drug-resistant HIV-1 variants. *Lancet* 1999; **354**: 729–33.
15. Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, Collier AC, *et al.* Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 2002; **347**: 385–94.
16. Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, Lalloo U, *et al.* Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet* 2006; **368**: 1575–80.
17. Guillebaud J. Any questions. *BMJ* 1993; **307**: 617.
18. Smith SK, Kirkman RJE, 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**: 513–22.
19. Sackett OL, Haynes RB, Gibson ES, Hackett BC, Taylor DW, Roberts RS, *et al.* Randomised clinical trial of strategies for improving medication compliance in primary hypertension. *Lancet* 1975; **1**: 1205–7.
20. Gross R, Bilker WB, Friedman HM, Strom BL. Effect of adherence to newly initiated antiretroviral therapy on plasma viral load. *AIDS* 2001; **15**: 2109–17.
21. Gross R, Bilker W, Friedman H, Coyne J, Strom B. Provider inaccuracy in assessing adherence and outcomes with newly initiated antiretroviral therapy. *AIDS* 2002; **16**: 1835–7.
22. Horwitz RI, Horwitz SM. Adherence to treatment and health outcomes. *Arch Intern Med* 1993; **153**: 1863–8.
23. Papasavvas E, Kostman JR, Mounzer K, Grant RM, Gross R, Gallo C, *et al.* Randomized, controlled trial of therapy interruption in chronic HIV-1 infection. *PLoS Med* 2004; **1**: e64.
24. Amico KR, Fisher WA, Cornman DH, Shuper PA, Redding CG, Konkle-Parker DJ, *et al.* Visual analog scale of ART adherence: association with 3-day self-report and adherence barriers. *J Acquir Immune Defic Syndr* 2006; **42**: 455–9.
25. 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 Acquir Immune Defic Syndr* 2007; **46**: 402–9.
26. Knobel H, Alonso J, Casado JL, Collazos J, Gonzalez J, Ruiz I, *et al.* Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study. *AIDS* 2006; **16**: 605–13.
27. Bangsberg DR, Bronstone A, Chesney MA, Hecht FM. Computer-assisted self-interviewing (CASI) to improve provider assessment of adherence in routine clinical practice. *J Acquir Immune Defic Syndr* 2002; **31**: S107–11.
28. Lu M, Safren SA, Skolnik PR, Rogers WH, Coady W, Hardy H, *et al.* Optimal recall period and response task for self-reported HIV medication adherence. *AIDS Behav* 2008: **12**: 86–94.
29. Simoni JM, Kurth AE, Pearson CR, Pantalone DW, Merrill JO, Frick PA. Self-report measures of antiretroviral therapy adherence: a review with recommendations for HIV research and clinical management. *AIDS Behav* 2006; **10**: 227–45.
30. Steiner JF, Koepsell TD, Fihn SD, Inui TS. A general method of compliance assessment using centralized pharmacy records: description and validation. *Med Care* 1988; **8**: 814–23.
31. Grossberg R, Zhang Y, Gross R. A time-to-prescription-refill measure of antiretroviral adherence predicted changes in viral load in HIV. *J Clin Epidemiol* 2004; **57**: 1107–10.
32. Low-Beer S, Yip B, O'Shaughnessy MV, Hogg RS, Montaner JS. Adherence to Triple Therapy and Viral Load Response. *J Acquir Immune Defic Syndr* 2000; **23**: 360–1.
33. Bangsberg DR, Hecht FM, Charlebois ED, Zolopa AR, Holodniy M, Sheiner L, *et al.* Adherence to protease inhibitors, HIV-1 viral load, and development of drug resistance in an indigent population. *AIDS* 2000; **14**: 357–66.
34. Bangsberg D, Perry S, Charlebois ED, Clark RA, Roberston M, Zolopa AR, *et al.* Non-adherence to highly active antiretroviral therapy predicts progression to AIDS. *AIDS* 2001; **15**: 1181–3.
35. Kalichman SC, Amaral CM, Stearns H, White D, Flanagan J, Pope H, *et al.* Adherence to antiretroviral therapy assessed by unannounced pill counts conducted by telephone. *J Gen Inter Med* 2007; **22**: 1003–6.

36. Toussi M, Choleau C, Reach G, Cahané MA. Novel method for measuring patients' adherence to insulin dosing guidelines: introducing indicators of adherence. *BMC Med Inform Decis Mak* 2008; **8**: 55.
37. Bernard L, Vuagnat A, Peytavin G, Hallouin MC, Bouhour D, Nguyen TH, *et al.* Relationship between levels of indinavir in hair and virologic response to highly active antiretroviral therapy. *Ann Intern Med* 2002; **37**: 656–9.
38. Gandhi M, Ameli N, Bacchetti P, Gange SJ, Anastos K, Levine A, *et al.* Protease inhibitor levels in hair strongly predict virologic response to treatment. *AIDS* 2009; **23**: 471–8.
39. Gandhi M, Greenblatt RM. Hair it is: the long and short of monitoring antiretroviral treatment. *Ann Intern Med* 2002; **137**: 696–7.
40. Gandhi M, Ameli N, Gange S, Anastos K, Bacchetti P, Cohen M, *et al.* Concentrations of efavirenz in hair correlate strongly with 24-hour intensive pharmacokinetic measurements and with virologic outcomes. *17th Conference on Retroviruses and Opportunistic Infections*, Feb 16–19, 2010.
41. Moore JD, Acosta EP, Johnson VA, Bassett R, Eron JJ, Fischl MA, *et al.* Intracellular nucleoside triphosphate concentrations in HIV-infected patients on dual nucleoside reverse transcriptase inhibitor therapy. *Antivir Ther* 2007; **12**: 981–6.
42. 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**: 497–504.
43. LoRe V, Amorosa VK, Localio AR, O'Flynn R, Teal V, Dorey-Stein Z, *et al.* Adherence to hepatitis C virus therapy and early virologic outcomes. *Clin Infect Dis* 2009; **48**: 186–93.
44. Donnelly L, Morris A, Evans J, and the DARTS/MEMO collaboration. Adherence to insulin and its association with glycaemic control in patients with type 2 diabetes. *Q J Med* 2007; **100**: 345–50.
45. Gazmararian JA, Kripalani S, Miller MJ, Miller MJ, Echt KV, Ren J, *et al.* Factors associated with medication refill adherence in cardiovascular-related diseases: a focus on health literacy. *J Gen Intern Med* 2006; **21**: 1215–21.
46. Taha TE, Kumwenda NI, Kafulafula G, Makanani B, Nkhoma C, Chen S, *et al.* Intermittent intravaginal antibiotic treatment of bacterial vaginosis in HIV-uninfected and -infected women: a randomized clinical trial. *PLoS Clin Trials* 2007; **2**: e10.
47. Wallace AR, Teitelbaum A, Wan L, Mulima MG, Guichard L, Skiler S, *et al.* Determining the feasibility of utilizing the microbicide applicator compliance assay for use in clinical trials. *Contraception* 2007; **76**: 53–6.
48. Hogarty K, Kasowitz A, Herold BC, Keller MJ. Assessment of adherence to product dosing in a pilot microbicide study. *Sex Transm Dis* 2007; **34**: 1000–3.
49. Julius SM, Sherman JM, Hendeles L. Accuracy of three electronic monitors for metered-dose inhalers. *Chest* 2002; **121**: 871–6.
50. Okeke CO, Quigley HA, Jampel HD, Ying G, Plyler RJ, Jiang Y, *et al.* Adherence with topical glaucoma medication monitored electronically the Travatan dosing aid study. *Ophthalmology* 2009; **116**: 191–9.
51. Albert JM, Demets DL. On a model-based approach to estimating efficacy in clinical trials. *Stat Med* 1994; **13**: 2323–35.
52. Coronary Drug Research Group. Influence of adherence to treatment and response of cholesterol on mortality on the coronary drug project. *N Engl J Med* 1980; **302**: 1038–41.
53. Lipid Research Clinics Program. The lipid research clinics coronary primary prevention trial results: Reduction in incidence of coronary heart disease *J Am Med Assoc* 1984; **251**: 351–64.
54. Acri T, Grossberg R, Gross R. How long is the right interval for assessing antiretroviral pharmacy refill adherence? *J Acquir Immune Defic Syndr* 2010; **54**: e16–e18.
55. Parienti JJ, Ragland K, Lucht F, de la Blanchardière A, Dargère S, Yazdanpanah Y, *et al.* Average adherence to boosted protease inhibitor therapy, rather than the pattern of missed doses, as a predictor of HIV RNA replication. *Clin Infect Dis* 2010; **50**: 1192–7.
56. Haynes RB, Taylor DW, Sackett DL, eds. *Compliance in Health Care*. Baltimore: Johns Hopkins University Press, 1979.
57. Feldman HI, Hackett M, Bilker W, Strom BL. Potential utility of electronic drug compliance monitoring in measures of adverse outcomes associated with immunosuppressive agents. *Pharmacoepidemol Drug Saf* 1999; **8**: 1–14.
58. Gardner EM, Sharma S, Peng G, Hullsiek KH, Burman WJ, MacArthur RD, *et al.* Differential adherence to combination antiretroviral therapy is associated with virological failure with resistance. *AIDS* 2008; **22**: 75–82.
59. Bangsberg DR, Deeks SG. Spending more to save more: interventions to promote adherence. *Ann Intern Med* 2010; **152**: 54–6.

60. University of Florida. Rx for health: engineers design pill that signals it has been swallowed. *Product Design and Development,* 2010, March 31. http://www.pddnet.com/news-rx-for-health-engineers-design-pill-that-signals-it-has-been-swallowed-033110/ (accessed July 2011).

61. Anonymous. Xhale's breathalyzer technology monitors patient drug compliance. *Med Gadget* 2008, April 22. http://medgadget.com/archives/2008/04/page/5 (accessed July 2011).

CHAPTER 43
Risk Evaluation and Communication

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

Introduction

All medications have risks. Some of these 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 health-care 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 health-care providers and health-care 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 determine that the medication is safe and effective (see Chapter 1).[1] Similar criteria are used by regulatory agencies in other countries as well.[2] 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.[3] 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 prescribers, procedures used to monitor potential adverse effects). 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.

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

Clinical problems to be addressed by pharmacoepidemiologic research

Although many different definitions exist, risk is usually defined as a potential harmful outcome that can occur with a known or unknown probability.[4] The two primary dimensions of risk are: (i) the probability with which the risk will occur, and (ii) the severity of the risk if it occurs. However, numerous subdimensions must also be evaluated.[5] For example, the probability that some risks will occur may be minimized by preventive actions (e.g., taking ulcerogenic medications with food) or early detection efforts (e.g., laboratory monitoring to detect signs of liver toxicity). Thus, they are controllable to some extent. Risks also differ in terms of whether potential harm caused is reversible or irreversible when detected and whether deleterious effects usually occur soon after initiation of therapy or may not arise for many years. In general, risks that are unlikely to occur for many years may evoke less concern than more proximal risks.[6]

Uncertainty is an inherent characteristic of any risk, but 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.[7] 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 (see also Chapters 4 and 10). Finally, the probability and severity of a particular risk is not necessarily invariant across different patient populations. For example, the risk of upper gastrointestinal bleeding associated with the use of non-steroidal anti-inflammatory medications increases with age, disproportionately affecting older adults.[8] Thus, to be able to make informed clinical judgments concerning the acceptability of specific medication risks, decision-makers (e.g., prescribers, patients, caregivers) must have sound, objective, scientific evidence concerning each of these various aspects of risk.

Evaluating whether a particular risk is worth taking also requires sound, objective, scientific evidence concerning treatment benefits and therapeutic alternatives. Serious risks associated with a particular medication may be acceptable if the medication offers substantial benefits, especially if no acceptable therapeutic alternatives are available.[6] However, the same risk may be unacceptable for a less effective medication that does not provide unique advantages over therapeutic alternatives.

Two broad categories of factors that can affect judgments concerning the acceptability of a risk have been identified: (i) characteristics of the adverse outcome and (ii) characteristics of the exposure.[6] Characteristics of the adverse outcome include: severity, frequency, reversibility, and predictability. In general, adverse events that are less serious, reversible, and are likely to affect only a small number of people, are likely to be judged more acceptable. Similarly, risks that are predictable and can be managed through appropriate care are likely to be judged as more acceptable. Other characteristics of the adverse outcome that can affect judgments of acceptability include whether the outcome is likely to occur immediately following exposure or be delayed for many years later and the level of certainty associated with the risk. In general, risks are less likely to be judged acceptable if they occur immediately following exposure and result from a well-established cause–effect relationship. Finally, certain types of outcomes (e.g., development of cancer) evoke more fear and emotional distress than other outcomes. Thus, risks associated with these types of outcomes are less likely to be considered acceptable.

As noted above, Strom also identified several characteristics of the exposure that can affect judgments of the acceptability of an adverse event. For example, risks of adverse events are likely to be judged more acceptable if a medication is essential to save or prolong life, no therapeutic alternatives available, and patients are fully informed of the risks associated with treatment so that they can decide whether or not to accept the associated risks. Finally, it is important to recognize that risk tolerance varies greatly across individuals. Thus, some

patients/providers may consider a certain risk acceptable, whereas others deem it unacceptable.

Collection of information concerning medication risks is pointless unless that information is made available to decision-makers in a timely manner and in a way that they can understand and use. Health-care providers have a professional responsibility to remain abreast of recent research findings. However, the sheer volume of emerging information can make this an onerous task. Conflicting findings from different studies adds to this burden.[9] Unfortunately, physicians often lack the skills in evidence-based medicine that they need to critically evaluate research findings.[10,11]

Health-care providers also have a responsibility to educate patients about the risks associated with medications prescribed for them, preventive actions patients should take to minimize these risks, warning signs patients should monitor while taking the medication, and actions patients should take if any warning signs appear. However, research suggests that the amount of information that physicians provide to patients regarding medication risks is limited. For example, in research involving arthritis patients, Katz and colleagues[12] found that rheumatologists stated the purpose for a newly prescribed non-steroidal anti-inflammatory drug in 91% of patient encounters, but mentioned medication risks in only 74% of encounters. The amount of information about medication risks provided was not assessed. In other areas, Scherwitz and colleagues[13] found that side effects were discussed in less than 50% of primary care physician–patient encounters, Rimer and colleagues[14] reported that oncologists disclosed only 69% of the information relevant to decisions concerning chemotherapy, and Richard and Lussier[15] found that potential adverse reactions were discussed in fewer than 17% of physician office visits in which a new medication was prescribed. In some cases, physicians may be reluctant to discuss possible medication risks with patients due to concern that it may decrease patient adherence to the prescribed medication regimen.[16] 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.[17–20]

Pharmacists also have a professional obligation to counsel patients about medication risks. However, few states in the United States require pharmacists to provide verbal counseling to patients when prescriptions are filled. Instead, most states require only that pharmacists offer to counsel patients.[21] 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.[22]

In the United States, most of the information that patients receive from pharmacists concerning medication risks is in the form of written materials that are distributed with prescriptions. However, a study reported in 2007 found that, although most pharmacies in the United States distribute written materials with prescription medications, many of the materials distributed failed to include information such as contraindications and precautions needed for safe medication use.[23] Notably, there was considerable variability in the written medication information distributed by pharmacies in the three countries examined, the United States, Australia, and the United Kingdom. The materials distributed in the United States were evaluated the least favorably.

Educating consumers about medication risks is also complicated by the amount of information about other types of risks that consumers are exposed to on a daily basis. Simply walking around one's home provides numerous examples. For example, a common label on kitchen step stools warns users that failure to read and follow all instructions may result in injury or death. These ever-present warnings, although perhaps beneficial in some respects, have the potential to desensitize consumers to information about serious medication risks. Moreover, written consumer medication information often appears to be designed to minimize liability, rather than to improve clinical

decision-making and patient health outcomes. Thus, many of these materials contain a long list of potential risks, but provide little useful information about any of them (e.g., probability of occurrence, potential impact if it occurs, how risk can be minimized).

In addition to the information about medication risks that is disseminated by health-care providers, consumers may obtain information from a wide variety of sources, including the Internet, direct-to-consumer advertising, and family and friends.[24,25–27] Unfortunately, the accuracy of available information varies widely from source to source, and few safeguards are in place to allow consumers to evaluate the quality of information available from different sources.

Finally, research also suggests that many patients may have difficulty understanding information concerning medication risks because of limited health literacy and numeracy skills.[28–30] Health literacy has been defined as "the degree to which individuals can obtain, process, and understand basic health information and services needed to make appropriate health decisions."[31] Numeracy, which involves the ability to understand and use numerical information, is one component of health literacy.[30] A systematic review of over 300 studies found that, on average, 26% of patients had inadequate health literacy skills and that an additional 20% had only marginal skills. Recent research suggests that many patients with limited health literacy can be identified using single-item screener questions.[28] The more difficult challenge is likely to involve developing risk communication strategies and tools that overcome the barriers posed by low health literacy and numeracy.

Methodologic problems to be addressed by pharmacoepidemiologic research

In addition to the clinical issues described above, many methodologic issues must be considered when communicating information about medication risks. First, one must determine which risks to communicate, the amount of detail to include in the communication, and the level of intensity with which to deliver the communication. Principles of informed consent, informed and shared decision-making, and professional ethics highlight the importance of patients understanding both the risks and benefits of different treatment options.[32–37] Moreover, research demonstrates that most patients want information about treatment risks.[38,39] For example, in a study of patients visiting outpatient clinics, Ziegler and colleagues[38] found that over 75% of patients wanted information about all possible adverse medication effects, no matter how rare, and that nearly 85% wanted information about all serious adverse effects, again no matter how rare. Fraenkel and colleagues[39] reported similar findings in a study involving rheumatoid arthritis (RA) patients. In that study, over 89% of study participants agreed with the statement: "It is important for me to know all the side effects of my medications."

Despite this interest in information about medication risks, research also suggests that patients often have a poor understanding of these risks. For example, a study conducted by Fraenkel and colleagues suggests that if patients were aware of the risks associated with many medications used to treat RA, they would not take them. In this study, RA patients were asked how willing they would be to take a medication that was associated with various side effects.[40] Willingness to take the medication was evaluated in relation to 17 different side effects, selected because they are associated with use of non-steroidal anti-inflammatory drugs (NSAIDs), low-dose prednisone, or the older disease-modifying antirheumatic drugs (DMARDs). Realistic probability information was provided. Although most of the patients were currently using at least one DMARD, 38% indicated that they would be unwilling, under any circumstances, to accept risks associated only with cosmetic changes (e.g., hirsutism, alopecia, weight gain, acne). Similarly, 45% said they would be unwilling to accept risks associated with major toxicities (e.g., pneumonia, liver damage, ulcers). Next, patients were asked to rate their willingness to take the medication as the probability of side-effect occurrence was decreased far below the actual risk.[41]

Only minor increases in patients' willingness to accept the risks presented were observed. These findings are especially striking because most patients in the study were taking the medications that they rejected in the hypothetical scenarios.

The findings summarized above suggest a need to better inform patients about medication risks. However, simply providing patients with written materials containing long lists of potential side effects, or providing this information verbally, does little to serve this goal. Thus, there is a need to prioritize the types of risk information to be communicated and identify the most appropriate targets for different communications (e.g., health-care providers, patients with a specific health problem or taking specific medications, consumers in general). Ideally, communications targeted toward patients would be tailored on the basis of patient characteristics such as: preferred role in medical decision making, tolerance for different types of medication risks, 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).

As discussed in the previous section, lack of health literacy and numeracy are significant barriers to effective risk communication. Most risk communications include probabilistic information, which even health-care providers can find difficult to interpret.[42–44] Several different numerical formats are used to express risk estimates. These include: absolute risk, absolute risk increase, relative risk, relative risk increase, odds ratios, and number needed to harm.[45] Many studies have demonstrated that the numerical format used to express risk information can have a substantial impact on judgment and decision making.[46,47] For example, read the two scenarios presented in Table 43.1 and think about how likely you would be to take the medications described. We have shown these scenarios to pharmacy students and practicing pharmacists in classroom settings, specifying that the hypothetical medication described in both scenarios is used to treat a chronic, painful, but not life-threatening, disorder. In both groups, considerably more participants indicated that they were more likely to take the medication under the circumstances described in Scenario A than those described in Scenario B. However, a careful perusal of the two scenarios reveals that they are identical. Scenario A simply expresses the risk in absolute terms, whereas Scenario B expresses the risk in relative terms. (One can use the information provided in Scenario A to calculate the relative risk of lymphoma associated with the medication described: (3.6/100 000 − 2.4/100 000) / (2.4/100 000) * 100% = 50%). The findings that we have obtained when using these scenarios as part of classroom exercises is consistent with a substantial body of literature that confirms that individuals tend to be less likely to accept risks when they are conveyed using relative risk formats as opposed to absolute risk formats.[47] The effect of format on judgment and decision making raises ethical issues because it suggests that risk communicators may manipulate the decisions that others reach following information exposure by varying the format in which information about risks are expressed. Experts recommend against providing risk information only in relative terms, isolated from baseline rates and other information that

Table 43.1 Absolute versus relative risk

A	Imagine that a medication that your doctor prescribed for you increased the risk of developing lymphoma within the next 5 years from 2.4 in 100 000 to 3.6 in 100 000. How likely is it that you would take the medication? Very likely Somewhat likely Neither likely nor unlikely Somewhat unlikely Very unlikely
B	Imagine that a medication that your doctor prescribed for you increased the risk of developing lymphoma within the next 5 years by 50%. How likely is it that you would take the medication? Very likely Somewhat likely Neither likely nor unlikely Somewhat unlikely Very unlikely

would contextualize the risk.[48] However, many questions remain concerning the optimal way to present numerical risk estimates.

Risk information may also be conveyed graphically. However, the issue regarding relative versus absolute risk cannot be avoided by using graphs. Figure 43.1 shows two simple graphs. The top panel emphasizes the difference in the relative risk of liver toxicity between two hypothetical medications. In this panel, only information about the cases of liver toxicity is presented visually. The size of the patient population in which these cases occurred is given only in the wording in the panel. In contrast, in the bottom panel, the size of the patient population is depicted by the length of the bar. Thus, this second panel emphasizes the difference in the absolute risk of liver toxicity between the two medications. As one would expect from the discussion of relative versus absolute risk above, individuals tend to be more risk averse when shown the graph emphasizing relative risk.[47]

More sophisticated graphs include pictographs, such as the one shown in Figure 43.2.[49,50] 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 5 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 5 years of therapy initiation. The dark grey 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).

Two other types of graphs are designed to facilitate communication regarding low-probability risks. First, Woloshin and colleagues have developed a risk magnifier scale that ranges from 0% (i.e., no chance that the risk will occur) to 100% (i.e., the risk is certain to occur).[47] Thus, it depicts the full range with which risks can occur. However, it enlarges the section of the scale between 0% and 1%, allowing for greater differentiation among low-probability risks. Second, the Paling Perspective Scale contextualizes low-probability risks by placing them on a scale that includes the probability of experiencing more familiar risks (e.g., the probability of being killed in an automobile crash).[46]

Considerable research has also focused on the development of standardized definitions that allow probabilistic information to be expressed in words (e.g., common, rare) rather than numbers. One set of verbal conventions range from "Very Common," to describe risks that affect greater than 10% of exposed patients, to "Very Rare" to describe risks that affect fewer than 1 in 10 000 exposed patients (i.e., 0.01%).[4,46] Another set of conventions range from "High," suggested for risks that affect more than 1% of exposed patients, to "Negligible," suggested for risks that affect fewer than 1 out of every 1 000 000 patients exposed (i.e., 0.0001%).[51] However, research examining how patients interpret these types of terms has found that patients tend to infer that the risks are much more likely than the standard definitions indicate.[4] For example, participants in one study estimated that, among patients taking a medication with a side effect described as "rare," 21% would experience the side effect. However, guidelines define "rare" risks as those that affect fewer than 1 in 1000 patients (i.e., 0.1%).[52] Thus, when communicating medication risks, verbal descriptors should either be avoided or defined explicitly in numerical terms as part of the risk communication.

As described above, the vast majority of research on communication concerning medication risks has focused on how probabilistic information is best conveyed. However, it is also important to investigate how best to communicate information about other risk dimensions (e.g., potential severity, controllability). The findings by Fraenkel and colleagues discussed earlier in this chapter suggest that patients may not place much weight on the estimated probability with which different potential risks may occur.[40] In that study, participants were asked how likely they would be to take a hypothetical medication given certain side effects. Participants' judgments changed little in response

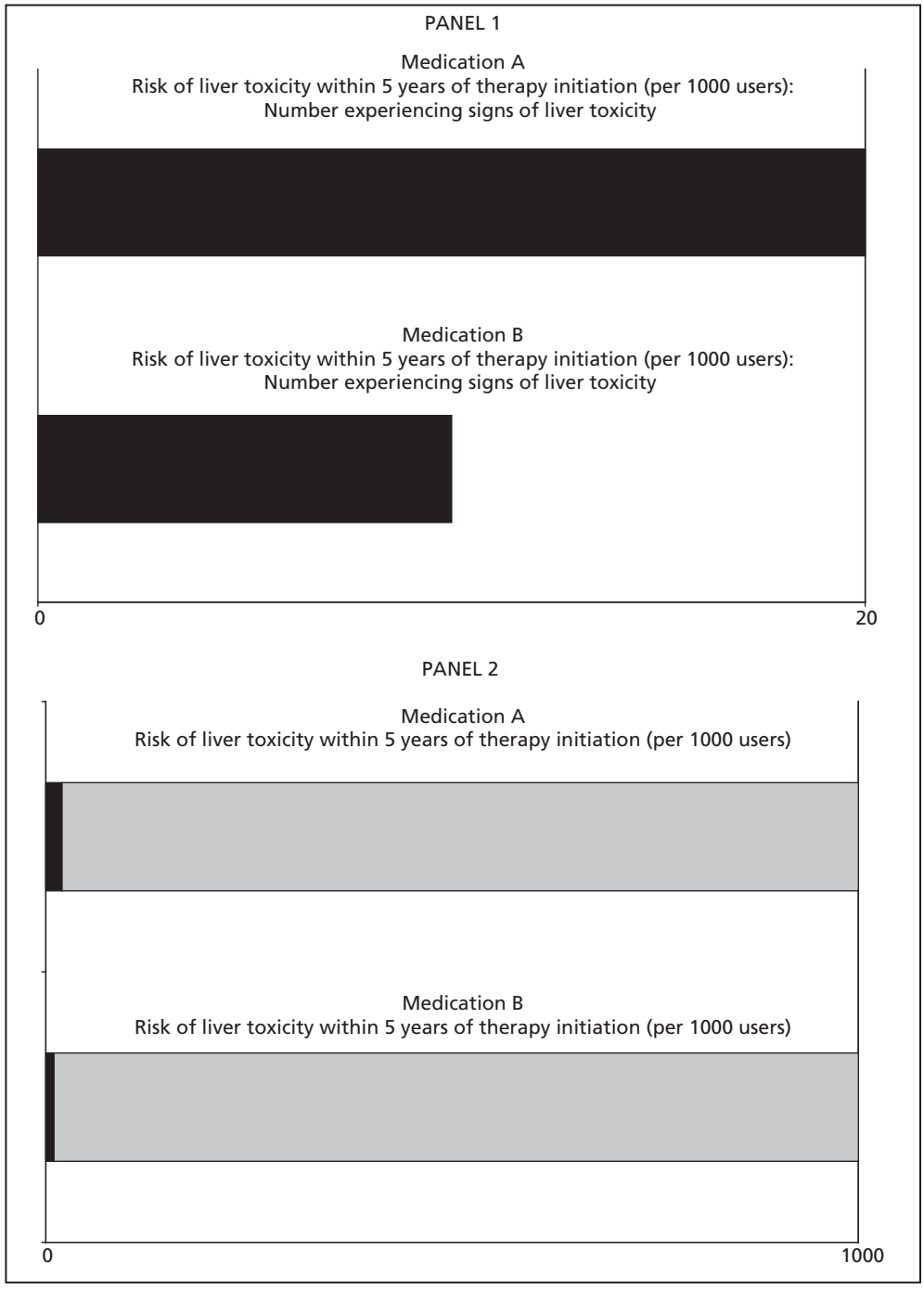


Figure 43.1 Figures emphasizing relative versus absolute risk. Black bar, people experiencing signs of liver toxicity; grey bar, people not experiencing liver toxicity.

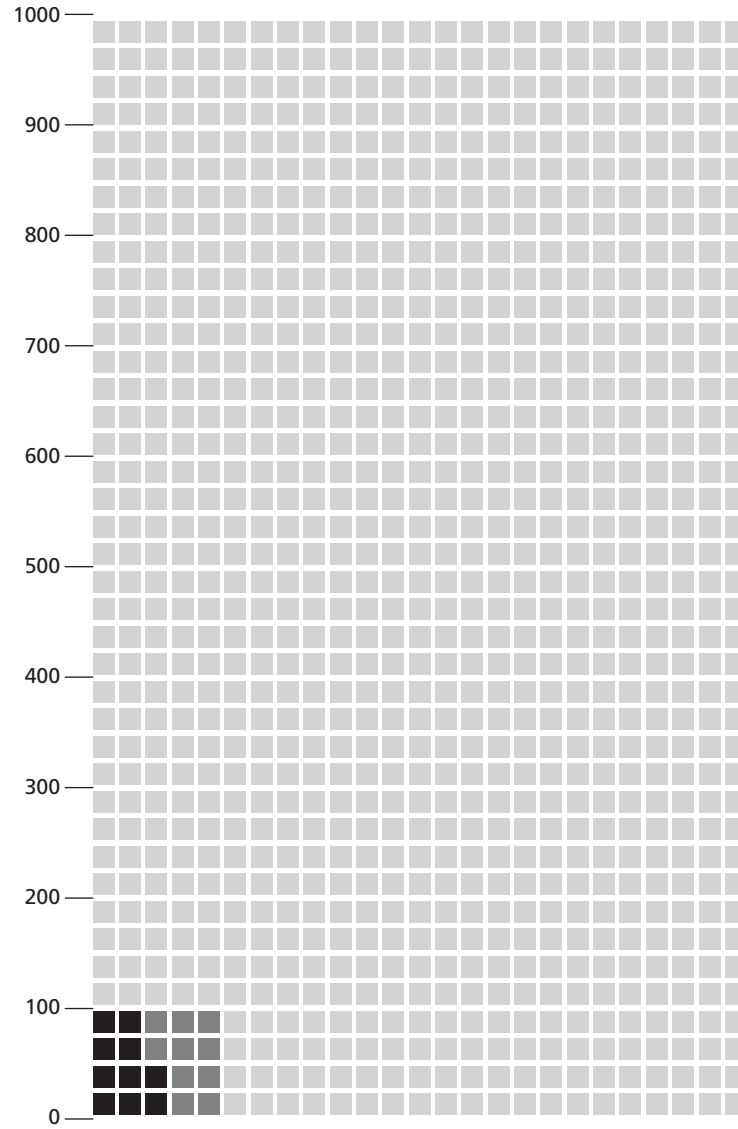


Figure 43.2 Sample pictograph.

to different probabilities. The investigators interpreted these findings as suggesting that patient judgments may be influenced more by the perceived impact that a potential side effect would have if it occurred, rather than on the probability that it will occur. The finding that patients who had experienced medication side effects previously were more likely to accept the risks associated with the hypothetical medication is consistent with this interpretation. These patients may have understood that many side effects are manageable with appropriate care. Conveying information concerning the potential impact different medication risks may have on health status and health-related quality of life is likely to be just as challenging as conveying probabilistic information. In fact, it may

be more challenging, because potential impact is inherently subjective and contingent upon a wide variety of factors. For example, medication risks that are associated with exposure to sunlight may have the greatest impact on those whose work requires them to be outside for extended periods of time during the day.

Finally, it is important to recognize that risk evaluation is not simply a cognitive exercise where estimates of probability and severity can be entered into a mathematical formula to derive an estimate of acceptability. It must also incorporate knowledge of how people respond affectively to risk information.[53–56] Research suggests that people tend to especially fear risks that have certain characteristics.[57] For example, risks that are poorly understood or are subject to contradictory statements from responsible sources tend to evoke greater fear than well-known risks. This is particularly relevant within the context of medication risk communication, where knowledge continues to evolve for years after a medication is introduced to the market (see Chapter 1). Consequently, patients may be exposed to a considerable amount of contradictory information over time. Experience with the selective cyclo-oxygenase-2 (COX-2) nonsteroidal anti-inflammatory medications and rosiglitazone provide ready examples. In addition, at any one point in time, different experts may have different opinions concerning a particular medication risk. Thus, a patient's physician might prescribe a medication for the patient and assure the patient that risks associated with the medication are minimal. However, the patient may read in the newspaper or on the Internet that other experts have concerns about the safety of the medication. This uncertainty may arouse significant anxiety, not only concerning the safety of the medication in question, but also concerning the safety of other, unrelated medications. That is, patients might ask themselves: "If experts don't know whether this medication is safe, how can they be sure about other medications." In addition, people tend to fear risks associated with man-made products more than those derived from natural sources. This tendency may explain, at least in part, why many patients fail to tell their health-care provider about the use of herbal remedies and natural products, and why many health-care providers fail to ask about their use.[58–61]

Other methodologic challenges concern how the effectiveness of risk communications are evaluated. 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 43.3, the most proximal effects of risk communications are likely to be increased knowledge and, in many cases, emotional arousal. The information communicated then may be incorporated into decision-making processes that lead to behavior change. However, if negative emotions are aroused (e.g., anxiety, fear, dread), message recipients will need to implement coping mechanisms that enable them to manage or control this emotional distress.[62–65] These coping mechanisms may either support or interfere with informed decision making. Thus, it is important for investigators

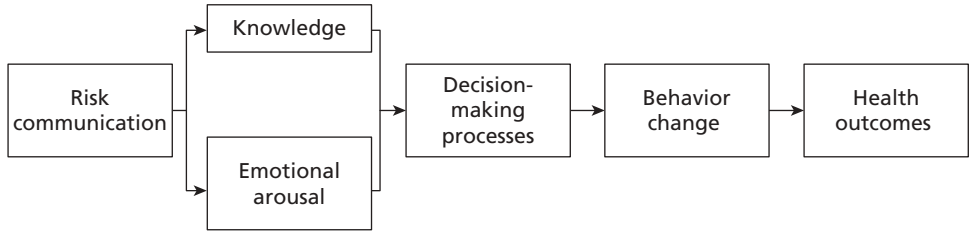


Figure 43.3 Conceptual model for evaluating the effectiveness of risk communication efforts.

to assess the emotional and affective effects that risk communications may have on message recipients.

It is also 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 behavior change. However, many risk communications include components that advocate specific actions (e.g., discontinuing a medication if a particular risk factor is present, initiating precautionary behaviors to reduce the 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. Thus, individuals developing risk communications should give careful consideration to the intended effects of the messages they develop and the time required for different types of effects to become evident.[66] For example, one would expect knowledge change to be evident immediately following message exposure. However, effects on health status are likely to require more time to appear.

Finally, it is important to consider unintended, as well as intended, effects. In addition, to causing emotional distress, 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.

Currently available solutions

Improving healthcare professional–patient communication about medication risks and benefits

Svarstad *et al*. sent trained shoppers who were acting as patients with new prescriptions into 306 community pharmacies in eight states.[67] The researchers defined any risk communication as providing information about one or more side effects or precautions. Adverse events were discussed for 17% of the new amoxicillin prescriptions, 31% of the new ibuprofen prescriptions, and 37% of the new paroxetine prescriptions. The researchers found patients who received prescriptions in states with more regulatory intensity surrounding pharmacist counseling (e.g., states that require that patients must be given face-to-face counseling by pharmacists) were more likely to receive risk information than patients in states with less regulatory intensity (e.g., states that only mandate that an offer for pharmacist counseling be given). State pharmacy boards need to consider requiring that patients must receive face-to-face counseling by pharmacists so that risk–benefit communication can be improved.

We were able to locate two randomized trials that successfully improved health-care professionals' risk communication skills.[68,69] In Wales, Elwyn and colleagues conducted a randomized trial that educated physicians about risk communication and shared decision making.[69,70] Physicians attended four workshops that were 3 hours each (two workshops were on risk communication and two were on shared decision making). The content of risk communication improved dramatically after the risk communication intervention, including the use of visual formats to help illustrate treatment risks to patients. Rickles *et al*. conducted a randomized, controlled trial where patients who were newly prescribed antidepressants received usual care from pharmacists or pharmacist-guided education and monitoring.[68] Patients were significantly more likely to report changes in depressive symptoms and side effects if they received pharmacist-guided education and monitoring. The study demonstrated that pharmacists can be trained to communicate better about antidepressant risks and benefits. Educational programs need to be developed to improve the risk communication skills of healthcare professionals.

Risk minimization action plans (RiskMAPs)

In the early part of this decade, several pharmaceuticals were removed from the market due to safety concerns. In reaction to this, the FDA formed three

working groups that discussed risk management. One result of these meetings was a draft guidance on RiskMAPs, which was issued in 2005.[71] The guidance gave manufacturers direction on how to develop objectives and goals as part of a risk minimization action plan to ensure that risks are minimized and benefits are maximized when pharmaceuticals are used. As an example, isotretinoin had a RiskMAP that had the goal of preventing use of the product in women of child-bearing age because of the drug's potential teratogenic effects.

Food and Drug Administration risk evaluation and mitigation strategies (REMS) (see also Chapter 29)

A 2006 Institute of Medicine report to the US Congress on drug safety criticized the FDA because of drug withdrawals from the market due to safety concerns.[72] As a result of this report and other events, the FDA Amendments Act (FDAAA -PL 110-85) of 2007 required companies to submit Risk Evaluation and Mitigation Strategies (REMS) to the FDA when they are necessary to ensure that the benefits of using a pharmaceutical outweigh the risks.[73] The FDA required that medications with RiskMAPs transition into REMS by September 2008. This caused RiskMAPs to become obsolete.

The FDA issued a draft guidance on REMS in September 2009.[74] The REMS must include product information and contact information for those responsible for the REMS. All REMS must have one or more overall goals. A proposed goal is the desired safety-related health outcome or the understanding by patients and/or health-care providers of the serious risk of a product. Examples of REMS goals include: "Patients taking drug X should be aware of the serious risks relative to the potential benefits," "Patients on drug X should not also take drug Y," or "Fetal exposure to drug X should not occur." Evaluations of REMS must examine whether the stated goals are achieved. This will be discussed in further detail below.

In some cases, the FDA requires companies to submit a REMS before approval and marketing of a drug. The FDA also has the power to require that companies submit a REMS after a drug has been approved if new safety information suggests that a REMS is needed to ensure that the drug's benefits outweigh the risks.[73] The following are possible sources of new safety information: adverse drug events (see Chapters 10, 19, and 20), the peer-reviewed biomedical literature, clinical trials, and the FDA's Sentinel Initiative.[75] The goal of the Sentinel Initiative is active surveillance of medication use to detect safety problems rather than passive surveillance (see Chapter 30).[76] Government and commercial databases are used as part of the initiative to conduct postmarketing surveillance to identify safety issues.

A proposed REMS may contain: (i) a Medication Guide and/or patient package insert and/or (ii) a communication plan targeted at health-care providers. It is important to point out that Medication Guides may be required for drugs that do or do not have a REMS. The industry is now required to add a toll-free number for reporting adverse events to all Medication Guides.[77] A communication plan might include: letters to health-care providers or disseminating information to providers through professional societies about any serious risks of a drug and any protocol to assure its safe use.

Elements to assure safe use may also be required as part of a REMS. These elements may include: patient–physician agreements or other informed consent procedures, patient education materials, safety protocols, medical or laboratory monitoring procedures, and data collection forms. Table 43.2

Table 43.2 Examples of elements to assure safe use

Health-care providers who prescribe the drug have particular training or experience, or are specially certified
Pharmacies, practitioners, or health-care settings that dispense the drug are specially certified
The drug be dispensed only in certain health-care settings, such as hospitals
The drug be dispensed to patients with evidence or other documentation of safe-use conditions, such as laboratory test results
Each patient using the drug to subject to certain monitoring
Each patient using the drug be enrolled in a registry

contains other elements that the FDA might require as part of a safe use plan.

In some REMS, companies are required to describe their implementation plan.[78] The FDA website contains a list of all approved REMS and the components included in each one (e.g., Medication Guide). For example, the REMS for ciprofloxacin tablets and oral suspensions is limited to a Medication Guide whereas the REMS for Epogen/Procit injection includes a MedGuide, a communication plan, elements to assure safe use, and an implementation plan.

Companies are also required to evaluate the effectiveness of their REMS strategies 18 months, 3 years, and 7 years after the strategy is approved. The evaluation results must be reported to the FDA so that it can be determined whether additional modifications to the REMS program are needed. Morris provides an excellent overview of four areas that REMS evaluators should consider when designing studies using surveys to assess the effectiveness of a REMS: (i) identifying the survey sample and administrating the survey, (ii) determining and justifying sample size, (iii) focusing survey questions on topics of need and interest, and (iv) designing questionnaires to minimize bias and provide useful information.[79] Morris points out that different risk management evaluations can lead to opposing viewpoints. He gives the example of troglitazone, a drug that can cause hepatotoxicity. The company distributed Dear Doctor letters and did a physician survey, which suggested that the majority of doctors were aware of the need for liver function testing and that the majority of patients received testing. In contrast, a FDA study using a managed care organization's database found that only 45% of patients were initially tested and only 2.5% were fully compliant with suggested liver function testing.[80] The FDA suggests that companies use at least two different types of study methods that complement each other's biases when evaluating their REMS. The FDA created a risk communication advisory committee in 2007. The committee advised the FDA to hire more behavioral scientists to evaluate the risk communication materials developed and used by pharmaceuticals companies. Much future work is needed to evaluate how effective pharmaceutical companies' risk communication methods are in educating patients about the risks and benefits of medications.

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 very much in its infancy. 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 potential outcomes. The conceptual model depicted in Figure 43.3 may help to structure future evaluation efforts.

More basic research is also needed to assess how people process and use information about medication risks. One promising approach involves the use of fuzzy-trace theory.[80–84] 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 pre-existing 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.[84]

Currently, much of the risk communication literature focuses on how probabilistic information is best conveyed. From this perspective, the difficulty patients have 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*.[85] 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: (i) how do patients 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, health-care providers, family/friends); (ii) what reasoning principles are invoked by contextual cues (e.g., format of the communication, images included in the communication) that affect patients' judgments and decisions concerning medication use; and (iii) what factors (e.g., limited health literacy skills, emotional state) interfere with information processing and lead to errors in reasoning.[82] 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.

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 health-care system. In addition, we must recognize that we live in the Information Age. Information about medications and medication risks is disseminated by many outside of the health-care 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. Meadows M. Promoting Safe and Effective Drugs for 100 Years. FDA Consumer Magazine, June 18, 2009. http://www.fda.gov/AboutFDA/WhatWeDo/History/ProductRegulation/PromotingSafeandEffectiveDrugsfor100Years/default.htm (accessed July 2011).
2. Lee PR, Herzstein J. International drug regulation. *Annu Rev Public Health* 1986; **7**: 217–35.
3. Lowrance WW. *Of Acceptable Risk*. Los Altos, CA: William Kaufmann, 1976.
4. Berry DC. *Risk, Communication and Health Psychology*. New York: Open University Press, 2004.
5. Bogardus ST, Jr, Holmboe E, Jekel JF. Perils, pitfalls, and possibilities in talking about medical risk. *JAMA* 1999; **281**: 1037–41.

6. Strom BL. When should one perform pharmacoepidemiology studies? In: Strom BL, ed. *Pharmacoepidemiology*, 2nd edn. New York: Wiley, 1994, pp. 59–66.
7. 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**: 2215–20.
8. Hernandez-Diaz S, Garcia-Rodriguez LA. Epidemiologic assessment of the safety of conventional nonsteroidal anti-inflammatory drugs. *Am J Med* 2001; **110** (Suppl 3A): 20S–7S.
9. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; **294**: 218–28.
10. Horton NJ, Switzer SS. Statistical methods in the journal. *N Engl J Med* 2005; **353**: 1977–9.
11. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA* 2007; **298**: 1010–22.
12. Katz JN, Daltroy LH, Brennan TA, Liang MH. Informed consent and the prescription of nonsteroidal antiinflammatory drugs. *Arthritis Rheum* 1992; **35**: 1257–63.
13. Scherwitz L, Hennrikus D, Yusim S, Lester J, Vallbona C. Physician communication to patients regarding medications. *Patient Educ Couns* 1985; **7**: 121–36.
14. Rimer B, Jones WL, Keintz MK, Catalano RB, Engstrom PF. Informed consent: A crucial step in cancer patient education. *Health Educ Q* 1984; **10** (Suppl): 30–42.
15. Richard C, Lussier MT. Nature and frequency of exchanges on medications during primary care encounters. *Patient Educ Couns* 2006; **64**: 207–16.
16. McGrath JM. Physicians' perspectives on communicating prescription drug information. *Qual Health Res* 1999; **9**: 731–45.
17. Bull SA, Hu XH, Hunkeler EM, Lee JY, Ming EE, Markson LE, *et al*. Discontinuation of use and switching of antidepressants: Influence of patient-physician communication. *JAMA* 2002; **288**: 1403–9.
18. 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**: 673–8.
19. Wilson SR, Strub P, Buist AS, Knowles SB, Lavori PW, Lapidus J, *et al*. Shared treatment decision making improves adherence and outcomes in poorly controlled asthma. *Am J Respir Crit Care Med* 2010; **181**: 566–77.
20. Horne R. Compliance, adherence, and concordance: Implications for asthma treatment. *Chest* 2006; **130** (1 Suppl): 65S–72S.
21. Resnik DB, Ranelli PL, Resnik SP. The conflict between ethics and business in community pharmacy: What about patient counseling? *J Bus Ethics* 2000; **28**: 179–86.
22. Puspitasari HP, Aslani P, Krass I. A review of counseling practices on prescription medicines in community pharmacies. *Res Social Adm Pharm* 2009; **5**: 197–210.
23. Raynor DK, Svarstad B, Knapp P, Aslani P, Rogers MB, Koo M, *et al*. Consumer medication information in the United States, Europe, and Australia: A comparative evaluation. *J Am Pharm Assoc (2003)* 2007; **47**: 717–24.
24. 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**: 565–70.
25. 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**: 41–9.
26. Narhi U, Helakorpi S. Sources of medicine information in Finland. *Health Policy* 2007; **84**: 51–7.
27. 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**: 169–76.
28. Powers BJ, Trinh JV, Bosworth HB. Can this patient read and understand written health information? *JAMA* 2010; **304**: 76–84.
29. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull* 2009; **135**: 943–73.
30. Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: Theory and practice. *Ann Behav Med* 2008; **35**: 261–74.
31. Institute of Medicine. *Health Literacy: a Prescription to End Confusion*. Washington DC: National Academies Press, 2004.
32. US President's Commission for the Study of Ethical Problems in Medicine and Biomedical And Behavioral Research. *Making Health Care Decisions: A Report on the Ethical and Legal Implications of Informed Consent in the Patient-Practitioner Relationship*. Washington, DC: US Government Printing Office, 1982.

33. Rimer BK, Briss PA, Zeller PK, Chan EC, Woolf SH. Informed decision making: What is its role in cancer screening? *Cancer* 2004; **101** (5 Suppl): 1214–28.
34. AMA. *AMA's code of Medical Ethics. E-10.01 Fundamental Elements of the Patient-Physician Relationship*, June 1992. http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics.page (accessed July 2011).
35. Whitney SN, McGuire AL, McCullough LB. A typology of shared decision making, informed consent, and simple consent. *Ann Intern Med* 2004; **140**: 54–9.
36. Charles C, Gafni A, Whelan T. Shared decision making in the medical encounter: What does it mean? (or, it takes at least two to tango). *Soc Sci Med* 1997; **44**: 681–92.
37. Charles C, Gafni A, Whelan T. Decision making in the physician-patient encounter: Revisiting the shared treatment decision making mode. *Soc Sci Med* 1999; **49**: 651–61.
38. Ziegler DK, Mosier MC, Buenaver M, Okuyemi K. How much information about adverse effects of medication do patients want from physicians? *Arch Intern Med* 2001; **161**: 706–13.
39. Fraenkel L, Bogardus S, Concato J, Felson D. Preference for disclosure of information among patients with rheumatoid arthritis. *Arthritis Rheum* 2001; **45**: 136–9.
40. Fraenkel L, Bogardus S, Concato J, Felson D. Unwillingness of rheumatoid arthritis patients to risk adverse effects. *Rheumatology (Oxford)* 2002; **41**: 253–61.
41. Fraenkel L, Bogardus S, Concato J, Felson D. Risk communication in rheumatoid arthritis. *J Rheumatol* 2003; **30**: 443–8.
42. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine. Communicating statistical information. *Science* 2000; **290**: 2261–2.
43. Monahan J. Statistical literacy: A prerequisite for evidence-based medicine. *Psychol Sci Public Interest* 2007; **8**: i–ii.
44. 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**: 53–96.
45. Crowson CS, Therneau TM, Matteson EL, Gabriel SE. Primer: Demystifying risk—understanding and communicating medical risks. *Nat Clin Pract Rheumatol* 2007; **3**: 181–7; quiz 1 p following 7.
46. Moore RA, Derry S, McQuay HJ, Paling J. What do we know about communicating risk? A brief review and suggestion for contextualizing serious, but rare, risk, and the example of cox-2 selective and non-selective NSAIDs. *Arthritis Res Ther* 2008; **10**: R20.
47. Woloshin S, Schwartz LM, Byram S, Fischhoff B, Welch HG. A new scale for assessing perceptions of chance: a validation study. *Med Decis Making* 2000; **20**: 298–307.
48. 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.
49. Zikmund-Fisher BJ, Fagerlin A, Roberts TR, Derry HA, Ubel PA. Alternate methods of framing information about medication side effects: Incremental risk versus total risk of occurrence. *J Health Commun* 2008; **13**: 107–24.
50. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. A demonstration of "less can be more" in risk graphics. *Med Decis Making* 2010; **30**: 661–71.
51. Calman KC, Royston GH. Risk language and dialects. *BMJ* 1997; **315**: 939–42.
52. Berry DC, Knapp P, Raynor DK. Provision of information about drug side-effects to patients. *Lancet* 2002; **359**: 853–4.
53. 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**: 1349–67.
54. Finucane ML. Emotion, affect, and risk communication with older adults: Challenges and opportunities. *J Risk Res* 2008; **11**: 983–97.
55. Slovic P, Peters E, Finucane ML, Macgregor DG. Affect, risk, and decision making. *Health Psychol* 2005; **24** (4 Suppl): S35–40.
56. 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**: 311–22.
57. Slovic P. Perception of risk. *Science* 1987; **236**: 280–5.
58. Eisenberg DM, Davis RB, Ettner SL, Appel S, Wilkey S, Van Rompay M, *et al*. Trends in alternative medicine use in the United States, 1990–1997: Results of a follow-up national survey. *JAMA* 1998; **280**: 1569–75.
59. Canter PH, Ernst E. Herbal supplement use by persons aged over 50 years in Britain: Frequently used herbs, concomitant use of herbs, nutritional supplements and prescription drugs, rate of informing doctors and

potential for negative interactions. *Drugs Aging* 2004; **21**: 597–605.

60. Gardiner P, Graham R, Legedza AT, Ahn AC, Eisenberg DM, Phillips RS. Factors associated with herbal therapy use by adults in the United States. *Altern Ther Health Med* 2007; **13**: 22–9.
61. Shelley BM, Sussman AL, Williams RL, Segal AR, Crabtree BF. 'They don't ask me so I don't tell them': Patient-clinician communication about traditional, complementary, and alternative medicine. *Ann Fam Med* 2009; **7**: 139–47.
62. Etchegary H, Perrier C. Information processing in the context of genetic risk: Implications for genetic-risk communication. *J Genet Couns* 2007; **16**: 419–32.
63. Gore TD, Bracken CC. Testing the theoretical design of a health risk message: Reexamining the major tenets of the extended parallel process model. *Health Educ Behav* 2005; **32**: 27–41.
64. McMahan S, Witte K, Meyer J. The perception of risk messages regarding electromagnetic fields: Extending the extended parallel process model to an unknown risk. *Health Commun* 1998; **10**: 247–59.
65. Leventhal H, Nerenz DR, Steele DS. Illness representations and coping with health threats. In: Taylor SE, Singer JE, eds. *Handbook of Psychology and Health*, Vol. 4. Hillsdale, NJ: Erlbaum, 1984.
66. 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.
67. Svarstad BL, 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.
68. Rickles NM, Svarstad BL, Stata-Paynter JL, Taylor LV, Kobak, KA. Improving patient feedback about and outcomes with antidepressant treatment: a study in eight community pharmacies. *J Am Pharm Assoc* 2006; **46**: 25–32.
69. Elwyn G, Edwards A, Hood K, Robling M, Atwell C. Achieving involvement: process outcomes from a cluster randomized trial of shared decision making skill development and use of risk communication aids in general practice. *Fam Pract* 2004; **21**: 337–46.
70. Edwards A, Elwyn G, Hood K, Atwell C, Robling M. Patient-based outcome results from a cluster randomized trial of shared decision making skill development and use of risk communication aids in general practice. *Fam Pract* 2004; **21**: 347–54.
71. Food and Drug Administration. *Guidance for Industry. Development and Use of Risk Minimization Action Plans*, March 2005. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126830.pdf (access July 2011).
72. Institute of Medicine. *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. Washington DC: National Academies Press, 2007.
73. 110th US Congress. *Food and Drug Administration Amendments Act of 2007*, September 27, 2007. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_public_laws&docid=f:publ085.110 (accessed July 2011).
74. Food and Drug Administration. *Guidance for Industry. Format and Content of Proposed Risk Evaluation and Mitigation Strategies (REMS), REMS Assessments, and Proposed REMS Modifications*, September 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM184128.pdf (accessed July 2011).
75. Meyer BM. The Food and Drug Administration Amendments Act of 2007: Drug safety and health-system pharmacy implications. *American Journal of Health Systems Pharmacy* 2009; **66** (Suppl 7): S3–5.
76. Shane R. Risk evaluation and mitigation strategies: Impact on patients, health care providers, and health systems. *Am J Health Syst Pharm* 2009; **66** (Suppl 7): S6–12.
77. Food and Drug Administration. *Guidance for industry. Medication Guides—Adding a Toll Free Number for Reporting Adverse Events*, June 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM164669.pdf (accessed July 2011).
78. Food and Drug Administration. *Approved Risk Evaluation and Mitigation Strategies (REMS)*, June, 2010. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111350.htm (accessed July 2011).
79. Morris L. Evaluating REMS tools: Issues in survey design. *Drug Inf J* 2009; **43**: 735–47.
80. Food and Drug Administration. *Endocrinologic and Metabolic Drug Advisory Committee Meeting. Briefing Book*. Slides part 1, March 1999. http://www.fda.gov/ohrms/dockets/ac/99/backgrd/3499b1.htm (accessed July 2011).
81. Reyna VF, Adam MB. Fuzzy-trace theory, risk communication, and product labeling in sexually transmitted diseases. *Risk Anal* 2003; **23**: 325–42.
82. Reyna VF. How people make decisions that involve risk: A dual-process approach. *Curr Dir Psychol Sci* 2004; **13**: 60–6.

83. 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, NJ: Lawrence Erlbaum Associates, 2005, pp. 77–106.
84. 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**: 179–95.
85. 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**: 203–11.

CHAPTER 44
Studying Effects of Antibiotics

Darren R. Linkin[1] *and Ebbing Lautenbach*[2]

[1]Division of Infectious Disease, Perelman School of Medicine at the University of Pennsylvania, and Philadelphia VA Medical Center, Philadelphia, PA, USA

[2]Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

Anti-infectives include antibiotics (or antibacterials), antivirals, and antifungal agents. A unique feature of many of these drugs, particularly the antibiotics, is that these agents affect not only the individual who consumes them, but also the larger microbial environment. As such, the impact of these drugs must be assessed both at the level of the individual as well as at the societal and environmental level. This chapter will focus primarily on antibiotics. Antibiotics are also unique in that they are auto-obsolete. With increased exposure to antibiotics, bacteria will, almost uniformly, elaborate mechanisms designed to evade or counteract these medicines. As such, the lifespan of use of an antibiotic is typically limited.

Despite these considerations, when antibiotics are prescribed, only the patient being treated is considered. While the care of the patient is certainly primary in this setting, a broader view of the impact of antibiotic use is needed. It is also important to note that the use and impact of antibiotics is not solely the purview of human medicine. Indeed, a large proportion of antibiotic use is found in other fields such as veterinary medicine, animal husbandry, and agriculture.

In light of these unique characteristics of antibiotics, it is not surprising that the most critical issue facing this class of drugs is the continued emergence of antibiotic resistance. The proportion of organisms demonstrating resistance to one or more antibiotics or antibiotic classes has increased exponentially in recent years. With increased global travel, dissemination of these resistant organisms to diverse parts of the world occurs ever more quickly. This has resulted in an increasingly larger number of organisms for which few, if any, antibiotics are available as therapeutic options. Reliance on costlier and more toxic agents (e.g., colistin, chloramphenicol) has resulted. Finally, this decrease in therapeutic options has coincided with a marked decrease in development of new antibiotics. In this climate, therapeutic options for the foreseeable future will be limited primarily to those antibiotics and antibiotic classes currently available. Preserving the utility of these agents is thus paramount.

The issue of antibiotic resistance is made all the more urgent when considered in the context of current trends in health-care-acquired infections. Health-care infections, particularly those due to resistant organisms, have increased markedly in recent years and present a tremendous challenge to the care of hospitalized patients. Strategies to limit the incidence and impact of these infections are urgently needed. These efforts will require both more rigorous evaluations of risk factors for infection with an antibiotic-resistant organism, as well as comprehensive assessments of the impact of intervention strategies.

The marked emergence of antibiotic resistance, combined with the lack of development of new

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

antibiotics, ushers in the very real possibility of a return to the preantibiotic era. Efforts to curtail the further emergence of resistance must focus on the conduct of sound epidemiologic research to better characterize the epidemiology of resistance and identify those strategies most likely to effectively counter current resistance trends. This chapter will focus on the clinical and methodologic impact of several emerging issues in pharmacoepidemiologic research, specifically the relationship between antibiotic use and antibiotic resistance.

Clinical problems to be addressed by pharmacoepidemiologic research

As noted above, the emergence of antibiotic resistance represents a growing crisis. The impact of resistance on the ability to treat bacterial infections has been profound, particularly given recent trends in the incidence of health-care-associated infections.

Infections are the most common adverse events encountered in health-care settings, with an estimated 5 to 10% of patients admitted to acute care hospitals becoming infected.[1] More than 80% of health-care-associated infections are caused by four types of infections: urinary tract infections, surgical site infections, bloodstream infections, and pneumonia.[2] Among identified pathogens in intensive care units (ICUs), 70% are resistant to at least one antibiotic.[3]

The emergence of antibiotic resistance has threatened to render the existing antibiotic arsenal useless. Both the number of organisms exhibiting resistance, as well as the mechanisms of resistance, have increased sharply in recent years.[4] The most common resistant Gram-positive organisms encountered in the health-care setting are methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant Enterococci (VRE). Among Gram-negative pathogens, extended-spectrum beta-lactamase (ESBL)-producing Enterobacteriaceae, multidrug-resistant *Pseudomonas aeruginosa*, and multidrug-resistant *Acinetobacter baumannii* are most prevalent.[2] As noted in data from the National Healthcare Safety Network (NHSN) of the Centers for Disease Control and Prevention (CDC), the prevalence of resistance for nearly all organisms studied increased substantially in recent years.[2]

Infections due to antibiotic-resistant bacteria are associated with increased morbidity, mortality, and hospital costs.[5–8] Studies focusing on the impact of specific antimicrobial resistant pathogens (e.g., MRSA, VRE, and ESBL-producing Enterobacteriaceae) have demonstrated significantly worse clinical outcomes in patients infected with resistant pathogens compared to patients infected with antibiotic-susceptible strains.[9–11]

The reason(s) for the apparent relationship between infection with an antibiotic-resistant organism and negative clinical outcomes has not been fully elucidated. One possible explanation is that antibiotic-resistant organisms are more virulent than their antibiotic-susceptible counterparts. However, the limited existing data suggest the opposite—that virulence factors and invasive disease are more common among antibiotic-susceptible strains.[12–14] Another possible explanation of the association between resistant infection and negative outcomes is that resistance may result in a delay in initiation of adequate antibiotic therapy (i.e., initiation of an antibiotic to which the organism is ultimately shown to be susceptible). Recent studies of bloodstream infections suggest that a delay in effective antibiotic therapy may result in poorer outcomes and that inadequate therapy is more likely to occur in resistant infections.[15–17] In a study of *Klebsiella* bacteremia, bloodstream infections that originated from bloodborne or respiratory infections had the highest mortality rate, whereas bacteremias that originated from a urinary tract infection had the lowest mortality rate.[18] Similarly, a recent cohort study of ESBL-producing Enterobacteriaceae noted that a delay in adequate antibiotic therapy was strongly associated with mortality in non-urine infections (e.g., blood, respiratory) but not in urine infections.[19] Finally, another possible explanation for the apparent association between antibiotic resistance

and negative clinical outcomes might be uncontrolled confounding. Patients with infections due to antibiotic-resistant organisms often have more comorbidities and a greater severity of illness than patients with infections due to antibiotic-susceptible organisms. While variables measuring underlying illness and other risk factors for poor outcomes are usually assessed and controlled for in studies seeking to examine the impact of resistance on outcomes, the possibility for residual confounding certainly exists.

The increasing trends in antibiotic resistance are of particular concern given marked slowing in development of new antibiotic agents.[20] A recent report noted a 75% decrease in systemic antibacterials approved by the Food and Drug Administration (FDA) from 1983 through 2007, with evidence of continued decrease in approvals, even during the most recently reported 5-year period (2003–2007).[21] Reports about the diminished discovery research efforts in large pharmaceutical companies and the decrease in antibacterial trials, most notably "early-phase" clinical trials, further highlight the ever lower industry focus on antibacterial drug research and development.[22,23] Only a handful of major pharmaceutical companies still have active antibacterial discovery programs, and the number of registered antibacterial trials decreased between 2005 and 2007.[22,23] Reasons for this shift away from antibacterial drug development include a greater focus on "lifestyle" drugs (e.g., sildenafil), emphasis on developing drugs for chronic illnesses for which medications must be taken for months or years (e.g., diabetes, hypertension), and ongoing controversies regarding FDA requirements for study sample size and endpoint definitions.[22,23]

The continued emergence of antibiotic resistance has been linked most closely to the selective pressure of the widespread use and misuse of antibiotic agents.[24,25] Under pressure of antibiotic use, the environmental microbial flora of medical care institutions increases in antibiotic resistance.[26–30] Patient acquisition of these resistant microbial organisms begins shortly after admission to a hospital and is accelerated by antibiotic treatment or prophylaxis.[31] This acquisition is paralleled by an increase in the prevalence of these bacteria as causative agents in health-care infections.

Given the close relationship between antibiotic use and antibiotic resistance, efforts to optimize antibiotic use are paramount. A greater understanding of precisely how antibiotics are used is a necessary first step. Antibiotics comprise the second most commonly used class of drugs in hospital formularies. Thirty to 65% of hospitalized patients receive antimicrobial agents,[32,33] and expenditures for these drugs may comprise 10–40% of the hospital pharmacy budget.[32,34] It is estimated that 37% (range 25–50%) of antibiotic use in hospitals is inappropriate,[35–38] and hospitals are often where patterns are set for outpatient practice. As is clear from these data, the opportunities to improve the use of antibiotics are vast.

Further complicating the development of new approaches to address antibiotic resistance is the fact that there are often marked differences across institutions with regard to the prevalence and epidemiology of particular resistant organisms. As such, it is critical that research focuses on elucidating the etiology for differences in epidemiology of resistance across sites. Ultimately, approaches targeted to the unique needs of a specific institution must be developed.[24,39]

Methodologic problems to be addressed by pharmacoepidemiologic research

Elucidating the association between antibiotic use and antibiotic resistance

As noted above, antibiotic use has been pinpointed as a major driver for the emergence of antibiotic resistance. While general overuse or misuse of antibiotics has been commonly reported, more refined assessment of specific associations between antibiotic use and antibiotic resistance are needed. Indeed, identifying specific antibiotics, or antibiotic use patterns, that may be most important in driving emerging resistance is critical. If such antibiotic use characteristics can be identified, strategies designed to optimize use could be designed and tested.

Measurement of antibiotic exposure

Despite the need for careful evaluation of the association between antibiotic use and antibiotic resistance, many inconsistencies exist in the methods employed to date. The approaches used to define prior antibiotic exposure have been noted to differ considerably across studies.[40] A systematic review of all studies investigating risk factors for extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella* species (ESBL-EK) was conducted to elucidate these issues.[41] This report described how included studies reported the extent of prior antibiotic use (e.g., exposure yes/no vs. duration of exposure) as well as the impact of using different methods on study conclusions.[41] Among the 25 included studies, prior antibiotic use was defined as a categorical variable in 18 studies, four studies defined prior antibiotic exposure as a continuous variable, and three studies included both a categorical and a continuous variable to describe prior antibiotic exposure. Only one paper provided an explicit justification for its choice of variable to describe prior antibiotic exposure. The authors then re-analyzed data from a previously published ESBL-EK risk factor study,[11] developing two separate multivariable models, one in which prior antibiotic use was described as a categorical variable (e.g., exposure yes/no) and one in which antibiotic use was described as a continuous variable (e.g., antibiotic days). Results of the two models using different methodologic approaches differed substantially. Specifically, third-generation cephalosporin use was a risk factor for ESBL-EK when antibiotic use was described as a continuous variable but not when antibiotic use was described as a categorical variable.[41]

These results strongly suggest that assessing prior antibiotic use as a categorical variable may mask significant associations between prior antibiotic use and resistance. For example, when the categorical variable is used, a subject who received only 1 day of an antibiotic is considered the same as a subject who received 30 days of the same antibiotic. However, the risk of resistance is almost certainly greater in the subject who received 30 days. Describing prior antibiotic use as a continuous variable allows for a more detailed characterization of the association between length of exposure and resistance. Indeed, it has been noted that the use of cutpoints can result in misinterpretation of data and that dichotomizing continuous variables reduces analytic power and makes it impossible to detect non-linear relationships.[42] Further, the relationship between prior antimicrobial use and resistance may not be linear (i.e., the risk of resistance may not increase at a constant rate with increasing antimicrobial exposure). The risk of resistance may not increase substantially until a certain amount of antimicrobial exposure has been attained (e.g., a "lower threshold"). A more precise characterization of this "lower threshold" would serve to better inform antibiotic use strategies.

Categorization of antibiotics

Another methodologic issue that focuses on defining prior antimicrobial use centers around how specific antibiotic agents are grouped. For example, antibiotic use could be classified by the agent (e.g., cefepime), class (e.g., cephalosporins), or spectrum of activity (e.g., Gram-negative). Although antibiotics are frequently grouped together in classes, individual agents within the class may differ substantially,[29] and such categorizations may mask important associations. A recent study explored these issues, again focusing on ESBL-EK as a model.[43] In a systematic review, 20 studies of risk factors for ESBL-EK that met inclusion criteria revealed tremendous variability in how prior antibiotic use was categorized. Categorization of prior antibiotic use was defined in terms of the specific agents, drug class, and often a combination of both. No study justified its choice of categorization method. There was also marked variability across studies with regard to which specific antibiotics or antibiotic classes were assessed as possible risk factors. A majority of the studies (n = 16) specifically investigated the use of beta-lactam antibiotics as risk factors for ESBL-EK. A variable number of studies also examined the association between use of other antibiotics and ESBL-EK infection: aminoglycosides (nine studies), fluoroquinolones (ten studies), and trimethoprim–sulfamethoxazole (seven studies). In a re-analysis of data from a prior study of risk factors for ESBL-EK,[11] two separate

multivariable models of risk factors for ESBL-EK were constructed: one with prior antibiotic use categorized by class and the other with prior antibiotic use categorized by spectrum of activity.[43] The results of these multivariable models differed substantially. Subsequent work reported similar findings when focusing on risk factors for carbapenem-resistant *Pseudomonas aeruginosa*.[44]

Time window of antibiotic exposure assessment

Another important issue is how remote antibiotic use is assessed. Specifically, when one investigates the association between antibiotic-resistant infection and prior antibiotic use, how far back should antibiotic use be assessed? A recent systematic review of studies investigating risk factors for ESBL-EK (noted above),[41] found that the time window during which antibiotic use was reviewed ranged from 48 hours to 1 year prior to the resistant infection. Furthermore, studies often did not explicitly state how far back in time prior antibiotic use was assessed.[41] One might assume that relatively recent antibiotic is more likely to lead to emergence of resistant bacteria, although very recent antibiotic use (e.g., within a few days prior to the resistant infection) may more often reflect early empiric therapy for what is ultimately identified as a resistant infection. The optimal time window during which to assess antibiotic use also almost certainly depends on the organism and the resistance mechanisms being studied.

Confounding and lack of outcome independence in studies of antibiotics

Studies comparing treatment with one or more antibiotics can evaluate a variety of outcomes including adverse and therapeutic effects. Assessing the comparative impact of different antibiotic exposures on subsequent colonization or infection with an antibiotic-resistant organism present numerous challenges. Two issues in antibiotic studies that are often threats to validity are confounding by indication and lack of independence of outcomes.

First, confounding by indication is often a problem of observational studies of both clinical outcomes and antimicrobial resistance[45] (see also Chapters 37 and 47). Studies are often done to determine if combination therapy or newer, often broader-spectra agents are more effective in achieving clinical or microbiological cures. Alternatively, studies may examine whether certain antimicrobial agents or regimens are more likely to lead to antimicrobial resistant colonization or infection. However, during usual clinical practice, patients more likely to fail therapy (e.g., with more comorbidities or higher severity of illness) are more likely to be given combination therapy with broader-spectrum agents. Thus, the indication for choosing antimicrobials—the severity or complexity of patients' illness—will often confound the relationship between a particular antibiotic/ regimen and the study outcome.

The effect of confounding (by indication) on study results can either be away from or towards the null hypothesis. In studies of cure of infection, the confounding is often towards the null. For instance, if a broader-spectrum antibiotic agent/ regimen (vs. a narrower agent/ regimen) truly prevents death from a particular infection/ pathogen, the observed effect may be that the broader regimen makes no difference (unless the benefit of the broader regimen is large), since patients at higher baseline risk of death may be more likely to receive it. Conversely, confounding in studies of the development of antibiotic resistance is often away from the null. Such studies often ask whether receiving a broader antibiotic agent/regimen is a risk factor for antibiotic resistant colonization or infection. Patients with a more complex/severe illness again are more likely to receive both the broader treatment and develop resistance.

Next, lack of independence of observations is a problem of both observational and interventional antibiotic studies evaluating risk factors, or preventive treatments, for colonization or infection with an antibiotic-resistant pathogen. "Colonization pressure," the daily average proportion of inpatients on a ward colonized with a particular pathogen, is a strong predictor of becoming colonized with the pathogen.[46] Thus, a risk factor may lead to more events if there is higher colonization pressure, given that as more colonization events occur, others become more likely.

Assessing the impact of antimicrobial use interventions

When convincing data emerge regarding the association between use of a particular antibiotic or antibiotic class and the emergence of a new antibiotic-resistant organism or resistance mechanism, the next logical step is to study the effect of an intervention designed to limit use of the implicated antibiotic. In studies of antibiotic resistance, there are unique issues with regard to choosing the most appropriate study design.

In general, a well-designed and adequately powered, randomized controlled trial provides the strongest evidence for or against the efficacy of an antibiotic use intervention. However, there are several reasons why a randomized controlled trial may not be feasible in the study of antibiotic use interventions. Randomizing individual patients to an antibiotic use intervention may not be a reasonable approach if person-to-person transmission of the antibiotic-resistant organism being studied occurs. In this case, individual patients are not independent, given the possibility of transmission of the resistant organism across patients. One might consider randomizing specific units or floors within one institution to receive the intervention. However, these units are not self-contained and patients and health-care workers frequently move from unit to unit. Thus, any effect of altered antibiotic use patterns on reduced transmission/acquisition of new resistant infections noted in the intervention units are likely to also result in some reduction in resistant infections in non-intervention areas (i.e., contamination). This would bias the results toward the null hypothesis (i.e., no effect of the intervention). Finally, when an intervention must be instituted rapidly in response to an emerging issue (e.g., an outbreak of an antibiotic-resistant organism), the first priority is to address and resolve the issue. In this case, it would be unethical to randomize an intervention across patient groups.

Given the above considerations, a well designed, quasi-experimental study offers a compelling alternative approach, and one that is frequently employed in investigations evaluating the impact of antibiotic use interventions.[47] The goal of this study design, which is also frequently referred to as a "before–after" or "pre–post intervention" study,[48,49] is to evaluate an intervention without using randomization. The most basic type of quasi-experimental study involves the collection of baseline data, the implementation of an intervention, and the collection of the same data following the intervention. Numerous variations that strengthen quasi-experimental studies exist and include: (i) institution of multiple pretests (i.e., collection of baseline data on more than one occasion); (ii) repeated interventions (i.e., instituting and removing the intervention sequentially); and (iii) inclusion of a control group (i.e., a group on which baseline and subsequent data is collected but on which no intervention is implemented).

While often employed in evaluations of interventions to curtail antibiotic resistance and/or hospital infections, critical evaluations of the advantages and disadvantages of quasi-experimental studies in these settings have only recently been performed.[47,50] A systematic review of four infectious diseases journals found that during a 2-year period, 73 articles focusing on infection control and/or antimicrobial resistance used a quasi-experimental study design.[47] Of these articles, only 12 (16%) used a control group, three (4%) provided justification for the use of the quasi-experimental study design, and 17 (23%) mentioned at least one of the potential limitations of such a design.[47] More attention has recently focused on increasing the quality of quasi-experimental study design and conduct to enhance the validity of conclusions drawn regarding effectiveness of interventions in the areas of infection control and antibiotic resistance.[47]

Potential limitations of quasi-experimental studies include regression to the mean, uncontrolled confounding, and maturation effects. Implementation of an intervention is often triggered in response to a rise in the rate above the norm, as is the case in an outbreak setting.[51] The principle of regression to the mean predicts that these elevated rates will tend to decline, even without intervention. This may result in the false conclusion that an effect is due to the intervention.[48,49]

Uncontrolled confounding is most likely to occur when variables other than the intervention

change over time or differ when comparing the pre- and postintervention periods.[48,49] This limitation can be addressed by measuring known confounders (e.g., hospital census, number of admissions) and controlling for them in analyses. However, not all confounders are known or easily measured (e.g., quality of medical and nursing care).

Finally, maturation effects are related to natural changes that patients experience with the passage of time.[48,49] In addition, there are cyclical trends (e.g., seasonal variation) that may be a threat to the validity of attributing an observed outcome to an intervention.

Control group selection in studies of antimicrobial resistance

Many studies have focused on identifying risk factors for antimicrobial resistance. The majority of these studies have been case–control studies. Control selection in case–control studies is critical in ensuring the validity of study results (see Chapter 3). Recent work has highlighted this issue of control group selection specifically for studies of antibiotic resistance.[40,52–55]

Historically, two types of control groups have been used in studies of antimicrobial resistant organisms.[40] The first type of control group is selected from patients who do not harbor the resistant pathogen. The second type of control group is selected from among subjects with a susceptible form of the infection. For example, in a study of risk factors for infection with ESBL-EK in hospitalized patients, the first type of control group would be selected from among the general hospitalized patient population while the second control group would be selected from among those patients with a non-ESBL-producing *E. coli* and *Klebsiella* species. The choice of control group should be based primarily on the clinical question being asked.

A limitation of using the first type of approach (i.e., using patients without infection as controls), is that, in addition to identifying risk factors for resistance, this approach also identifies risk factors for infection with that organism in general (regardless of whether the infection is resistant or susceptible). Thus, there is no way to distinguish the degree to which a risk factor is associated with the resistance phenotype versus associate with the infecting organism in general.[53]

While use of the second type of control group (i.e., patients with a susceptible form of the infection) has historically been a more common approach, it has recently been demonstrated that use of this type of control group may result in an overestimate of the association between antimicrobial exposure and resistant infection.[54,55] Using the example of ESBL-EK above, the explanation for this finding has been postulated as follows: if the controls are represented by patients with non-ESBL-producing *E. coli* and *Klebsiella* infections, it is very unlikely that these patients would have recently received ceftazidime (i.e., the risk factor of interest) since exposure to ceftazidime would have eradicated colonization with non-ESBL-producing *E. coli* and *Klebsiella* infections. Thus, the association between ceftazidime use and ESBL-EK would be overestimated.[56]

Another concern with using the second type of control group (i.e., patients with a susceptible form of the infection) is the potential for misclassification bias. Specifically, subjects selected as controls who have never had a clinical culture obtained may in fact harbor unrecognized colonization with the resistant organism under study.[52] Since it is probable that patients colonized with the resistant organism would likely have had greater prior antimicrobial exposure than subjects not colonized, this misclassification would likely result in a bias toward the null (i.e., the cases and controls would appear falsely similar with regard to prior antimicrobial use). An additional concern with using the second type of control group and identifying as controls those patients who have never had a clinical culture, is that differences between cases and controls may simply reflect the fact that clinical cultures were performed for case patients but not for controls. Since procurement of cultures is not a random process but based on clinical characteristics, it is possible that the severity of illness or antibiotic exposure may be greater among cases, regardless of the presence of antibiotic resistant infection.[40] One potential approach would be to limit eligible controls to those patients for whom at

least one clinical culture has been performed and does not reveal the resistant organism of interest. Such a negative culture would suggest that the patient is likely not colonized with the resistant organism. However, recent work has demonstrated that using clinical cultures to identify eligible controls leads to the selection of a control group with a higher co-morbidity score and greater exposure to antibiotics compared with a control group for which clinical cultures were not performed.[52]

Currently available solutions

Outcomes in antimicrobial studies

One may minimize confounding by indication by carefully considering how subjects were given particular antibiotic exposures. Using subjects who were given either a study antibiotic/ regimen or the comparator without regard to their risk of the outcome will avoid confounding by indication. For example, a quasi-experimental study design may be used to examine a change from cefepime to piperacillin–tazobactam as the first-line broad-spectrum Gram-negative coverage for ventilator-associated pneumonia in an intensive care unit. Specifically, following a period of time during which cefepime was used as the first-line agent (i.e., the "pre" period), the first-line therapy would then be changed to piperacillin–tazobactam (i.e., the "post" period). The indication (e.g., severity of illness, risk of treatment failure) does not inform the antibiotic choice, avoiding confounding by indication. In this example, other confounding variables that may change over time (e.g., severity of illness, underlying disease) may of course remain of concern. As another example, in a cohort study, subjects with methicillin-susceptible *Staphylococcus aureus* bloodstream infections are sometimes treated with vancomycin instead of the preferred nafcillin or cefazolin due to a penicillin allergy. Since allergy (the indication for giving one therapy instead of another) is unlikely to be associated with the study outcome (i.e., cure of infection), this careful selection of exposed subject groups again avoids confounding by indication. Finally, one may carefully assess variables (e.g., severity of illness, underlying disease) that may lead to preferential selection of a given therapeutic option. In this approach, one may then try to control for these variables in various ways (e.g., logistic regression, propensity score analysis) when comparing outcomes of interest across the two therapies.[57,58]

As discussed previously, lack of independence of individual outcomes is often a hazard when using antibiotic resistance as an outcome. For studies examining the impact of different antibiotic regimens for antibiotic resistance outcomes (colonization or infection with antibiotic resistant pathogens), the "contamination" is literally cross-contamination or spread of the pathogens from patient to patient (e.g., via health-care workers' hands and other vectors). Cluster randomization can be done in which groups or centers rather than individuals are randomized, in order to minimize contamination between subjects randomized to different study arms. When the unit of randomization is a hospital (or even a unit), subjects in different study arms are geographically separated, decreasing the chance of cross-contamination between them. However, contamination is still possible if the health-care workers or equipment that serve as vectors travel between the hospitals or units. A similar effect (though without the strength of randomization) is using a quasi-experimental study design to evaluate different antibiotic regimens in different hospitals/ units, typically with fewer "centers" than a cluster randomized study.

Another approach for addressing contamination is to assess directly colonization with resistant organisms in different groups. If the colonizing/ infecting pathogens are prospectively collected by the study team, or saved from the clinical microbiology laboratory that processes clinical samples, clonality analyses can be performed. These analyses, using a number of potential methods (most commonly pulsed field gel electrophoresis) can ascertain the genetic relatedness of isolates, and thus their likelihood of representing person-to-person spread. To demonstrate development of resistance and/or emergence of a resistant clone in an individual subject, multiple surveillance cultures can be analyzed sequentially to determine evolution of resistance over time.

Assessing the impact of antimicrobial use interventions

As noted earlier, the quasi-experimental study design offers a valuable approach in investigating the impact of antibiotic use interventions. Several important limitations of this study design were noted earlier and include regression to the mean, uncontrolled confounding, and maturation effects. Several approaches have been employed to address these limitations. First, when addressing regression to the mean, incorporating a prolonged baseline period (i.e., a long "pre" period) prior to the intervention allows one to evaluate the natural fluctuation in rates of the outcome over time and permits a more comprehensive assessment of possible regression to the mean. Second, changes in the outcome of interest may be measured at a control site during the same time period. Finally, the use of segmented regression analysis may assist in addressing possible regression to the mean in that not only will the immediate change in prevalence coincident with the intervention be assessed, but also the change in slope over time.[59–61]

With regard to uncontrolled confounding, this limitation can be addressed by measuring known confounders (e.g., hospital census, number of admissions) and controlling for them in analyses. However, not all confounders are known or easily measured (e.g., quality of medical and nursing care). To address this, one may assess a non-equivalent dependent variable to evaluate the possibility that factors other than the intervention influenced the outcome.[47,50] A non-equivalent dependent variable is defined as a variable that has similar potential causal and confounding variables as the primary dependent variable except for the effect of the intervention. For example, in assessing the impact of an intervention to limit ceftazidime use on ESBL-EK prevalence, one might consider incidence of methicillin-resistant *Staphylococcus aureus* (MRSA) as a non-equivalent dependent variable. While ESBL-EK and MRSA might both be affected by such factors as the proportion of hospital beds filled, nurse-to-patient ratio, and infection control practices, it is unlikely that ceftazidime use specifically would affect the incidence of MRSA.

Maturation effects are related to natural changes that patients experience with the passage of time.[48,49] In addition, there are cyclical trends (e.g., seasonal variation) that may be a threat to the validity of attributing an observed outcome to an intervention. This potential limitation may be addressed through approaches noted above including assessment of a prolonged baseline period, use of control sites, implementing interventions at different time periods at different sites, and assessing a non-equivalent dependent variable.

Finally, recent concerted efforts to improve the conduct of quasi-experimental studies have been made. For example, the recent ORION (Outbreak Reports and Intervention Studies Of Nosocomial infection) statement presents a clear framework for how quasi-experimental studies are conducted and reported.[62] Future work will be required to determine the impact of these broader efforts to improve the quality of these studies.

Case and control group selection in studies of antimicrobial resistance

One proposed approach to addressing the difficulties in control group selection in studies of antimicrobial resistance is the case–case–control study design.[53,63–65] In this design, effectively two case–control studies are performed. In the first, cases are defined as those patients harboring the resistant organism while controls are those patients without the pathogen of interest. In the second, cases are instead defined as those patients harboring the susceptible bacteria while controls, similar to the first approach are those patients without the pathogen of interest.[53] These two separate studies are then carried out with risk factors from the two studies compared qualitatively. This approach allows for the comparison of risk factors identified from the two studies to indicate the relative contribution of the resistant infection over and above simply having the susceptible infection. A potential limitation in this approach is the difficulty in matching for potential confounders because of the use of only one control group.[53] Since there are two different case groups, case variables (e.g., duration of hospitalization, patient location) cannot be used for matching. In addition, the qualitative comparison of results

from the two studies in this design leaves open the question as to how much of a difference in results is meaningful. Further work in evaluating quantitative approaches to comparing the respective models created in such studies would be desirable.

The future

This chapter concentrates on several emerging issues in pharmacoepidemiologic research focused on the relationship between antibiotic use and antibiotic resistance. This is by no means an exhaustive review of all such methodologic issues in this area. Indeed, antibiotic resistance research is characterized more by what is not known, than what is known.

Much greater emphasis must be placed on determining the optimal use of antibiotics. For example, with few exceptions (e.g., endocarditis, osteomyelitis), the necessary duration of antibiotics for various infectious diseases is unknown. The ability to pinpoint more clearly the shortest effective duration of antibiotic therapy for various infections would be very beneficial in providing evidence-based recommendations for shortening current courses of antibiotic therapy. This major gap in knowledge has not gone unnoticed. Recent NIH initiatives have solicited randomized controlled trials comparing longer versus shorter durations of therapy for common infections. In addition, these recent initiatives have focused on comparing antibiotic treatment versus no treatment for those infections for which bacterial causes are very uncommon (e.g., upper respiratory tract infections). Future sustained emphasis on expanding the knowledge base regarding optimal use of antibiotics will be critical.

As noted previously, the epidemiology of a given antibiotic-resistant organism may vary across geographic regions and even across institutions within the same region. Antibiotic prescribing practices also vary substantially across centers. As such, there is a clear need for larger, multicenter studies of antibiotic resistance and interventions to improve antibiotic use. The need for cluster randomized trials is particularly evident given that certain interventions (e.g., changes in a hospital's antibiotic formulary) can only be tested at the institutional level. Therefore, the need for large networks working collaboratively to address these problems is clear. Indeed, recent efforts on the part of the CDC and the Society for Healthcare Epidemiology of America (SHEA) have begun to lay the groundwork for establishing such collaborative networks.

Finally, much of the research on antibiotic resistance conducted to date has focused on the hospital setting. The reasons for this are understandable because antibiotic resistant organisms typically emerge first in the hospital setting, transmission of resistant organisms is greater when patients are in close proximity, and studying antibiotic use is more straightforward in the hospital, particularly when investigators have the potential to alter the antibiotic formulary. Despite these factors, it is important to recognize that antibiotic resistance occurs in many other health-care and non-health-care settings. Community settings confront organisms distinct from the health-care setting (e.g., penicillin-resistant *Streptococcus pneumoniae*) as well as organisms that overlap considerably with health-care sites (e.g., methicillin-resistant *S. aureus*). Non-hospital health-care sites (e.g., long-term-care facilities, rehabilitation centers) face antibiotic resistant organisms that are as worrisome, if not more so, than hospital sites. In fact, the prevalence of antibiotic resistance in these sites has for the most part surpassed that of hospitals. Finally, it is well known that patients often move through various health-care settings (e.g., from a long-term-care facility to a hospital, then to a rehabilitation center). Such movement demonstrates that effective interventions to curb antibiotic resistance must account for the relationships between facilities in the greater health-care community, rather than treating each center in isolation. Only through such coordinated efforts can we learn what is required to successfully address the emerging crisis of antibiotic resistance.

References

1. Burke JP. Infection control—a problem for patient safety. *N Engl J Med* 2003; **348**: 651–6.

2. NNIS. National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004. *Am J Infect Control* 2004; **32**: 470–85.
3. Eggimann P, Pittet D. Infection control in the ICU. *Chest* 2001; **120**: 2059–63.
4. Tenover FC. Development and spread of bacterial resistance to antimicrobial agents: an overview. *Clin Infect Dis* 2001; **33** (Suppl 3): 108–15.
5. Cohen ML. Epidemiology of drug resistance: implications for a post-antimicrobial era. *Science* 1992; **257**: 1050–5.
6. Holmberg SD, Solomon SL, Blake PA. Health and economic impacts of antimicrobial resistance. *Rev Infect Dis* 1987; **9**: 1065–78.
7. Cosgrove SE. The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs. *Clin Infect Dis* 2006; **42** (Suppl 2): S82–9.
8. Cosgrove SE, Carmeli Y. The impact of antimicrobial resistance on health and economic outcomes. *Clin Infect Dis* 2003; **36**: 1433–7.
9. Vergis EN, Hayden MK, Chow JW, Snydman DR, Zervos MJ, Linden PK, *et al*. Determinants of vancomycin resistance and mortality rates in enterococcal bacteremia. A prospective multicenter study. *Ann Intern Med* 2001; **135**: 484–92.
10. Blot SI, Vandewoude KH, Hoste EA, Colardyn FA. Outcome and attributable mortality in critically Ill patients with bacteremia involving methicillin-susceptible and methicillin-resistant *Staphylococcus aureus*. *Arch Intern Med* 2002; **162**: 2229–35.
11. Lautenbach E, Patel JB, Bilker WB, Edelstein PH, Fishman NO. Extended-spectrum B-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*: risk factors for infection and impact of resistance on outcomes. *Clin Infect Dis* 2001; **32**: 1162–71.
12. Vila J, Simon K, Ruiz J, Horcajada JP, Velasco M, Barranco M, *et al*. Are quinolone-resistant uropathogenic *Escherichia coli* less virulent? *J Infect Dis* 2002; **186**: 1039–42.
13. Blazquez R, Menasalvas A, Carpena I, Ramirez C, Guerrero C, Moreno S. Invasive disease caused by ciprofloxacin-resistant uropathogenic *Escherichia coli*. *Eur J Clin Microbiol Infect Dis* 1999; **18**: 503–5.
14. Maslow JN, Lautenbach E, Bilker WB, Johnson JR. Colonization with extraintestinal pathogenic *Escherichia coli* among nursing home residents and its relationship to fluoroquinolone resistance. *Antimicrob Agents Chemother* 2004; **48**: 3618–20.
15. Kollef MH, Ward S, Sherman G, Prentice D, Schaiff R, Huey W, *et al*. Inadequate treatment of nosocomial infections is associated with certain empiric antibiotic choices. *Crit Care Med* 2000; **28**: 3456–64.
16. Lodise TP, McKinnon PS, Swiderski L, Rybak MJ. Outcomes analysis of delayed antibiotic treatment for hospital-acquired *Staphylococcus aureus* bacteremia. *Clin Infect Dis* 2003; **36**: 1418–23.
17. Harbarth S, Garbino J, Pugin J, Romand JA, Lew D, Pittet D. Inappropriate initial antimicrobial therapy and its effect on survival in a clinical trial of immunomodulating therapy for severe sepsis. *Am J Med* 2003; **115**: 529–35.
18. Paterson DL. "Collateral damage" from cephalosporin or quinolone antibiotic therapy. *Clin Infect Dis* 2004; **38** (Suppl 4): S341–5.
19. Hyle EP, Lipworth AD, Zaoutis TE, Nachamkin I, Bilker WB, Lautenbach E. Impact of inadequate initial antimicrobial therapy on mortality in infections due to extended-spectrum beta-lactamase-producing enterobacteriaceae: variability by site of infection. *Arch Intern Med* 2005; **165**: 1375–80.
20. Spellberg B, Powers JH, Brass EP, Miller LG, Edwards JE Jr. Trends in antimicrobial drug development: implications for the future. *Clin Infect Dis* 2004; **38**: 1279–86.
21. Spellberg B, Guidos R, Gilbert D, Bradley J, Boucher HW, Scheld WM, *et al*. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clin Infect Dis* 2008; **46**: 155–64.
22. Taubes G. The bacteria fight back. *Science* 2008; **321**: 356–61.
23. Karlberg JPE. Trends in disease focus of drug development. *Nat Rev Drug Discov* 2008; **7**: 639–40.
24. Muto CA, Jernigan JA, Ostrowsky BE, Richet HM, Jarvis WR, Boyce JM, *et al*. SHEA guideline for preventing nosocomial transmission of multidrug-resistant strains of *Staphylococcus aureus* and enterococcus. *Infect Control Hosp Epidemiol* 2003; **24**: 362–86.
25. Safdar N, Maki DG. The commonality of risk factors for nosocomial colonization and infection with antimicrobial-resistant Staphylococcus aureus, enterococcus, gram-negative bacilli, *Clostridium difficile*, and *Candida*. *Ann Intern Med* 2002; **136**: 834–44.
26. Emori TG, Gaynes RP. An overview of nosocomial infections, including the role of the microbiology laboratory. *Clin Microbiol Rev* 1993; **6**: 428–42.
27. McGowan JE, Jr. Antimicrobial resistance in hospital organisms and its relation to antibiotic use. *Rev Infect Dis* 1983; **5**: 1033–48.

28. Spach DH, Silverstein FE, Stamm WE. Transmission of infection by gastrointestinal endoscopy and bronchoscopy. *Ann Intern Med* 1993; **118**: 117–28.
29. Donskey CJ. The role of the intestinal tract as a reservoir and source for transmission of nosocomial pathogens. *Clin Infect Dis* 2004; **39**: 219–26.
30. Donskey CJ, Chowdhry TK, Hecker MT, Hoyen CK, Hanrahan JA, Hujer AM, *et al*. Effect of antibiotic therapy on the density of vancomycin-resistant enterococci in the stool of colonized patients. *N Engl J Med* 2000; **343**: 1925–32.
31. Selden R, Lee S, Wang WL, Bennett JV, Eickhoff TC. Nosocomial klebsiella infections: intestinal colonization as a reservoir. *Ann Intern Med* 1971; **74**: 657–64.
32. Bryan CS. Strategies to improve antibiotic use. *Infect Dis Clin North Am* 1989; **3**: 723–34.
33. Pakyz AL, MacDougall C, Oinonen M, Polk RE. Trends in antibacterial use in US academic health centers; 2002 to 2006. *Arch Intern Med* 2008; **168**: 2254–60.
34. Salama S, Rotstein C, Mandell L. A multidisciplinary hospital-based antimicrobial use program: Impact on hospital pharmacy expenditures and drug use. *Can J Infect Dis* 1996; **7**: 104–9.
35. Willemson I, Groenhuijzen A, Bogaers D, Stuurman A, van Keulen P, Kluytmans J. Appropriateness of antimicrobial therapy measured by repeated prevalence surveys. *Antimicrob Agents Chemother* 2007; **51**: 864–7.
36. Maki DG, Schuna A. A study of antimicrobial misuse in a university hospital. *Am J Med Sci* 1987; **275**: 271–82.
37. Hecker MT, Aron DC, Patel NP, Lehmann MK, Donskey CJ. Unnecessary use of antimicrobials in hospitalized patients: current patterns of misuse with an emphasis on the antianaerobic spectrum of activity. *Arch Intern Med* 2003; **163**: 972–8.
38. Dunagan WC, Woodward RS, Medoff G, Gray JL 3rd, Casabar E, Smith MD, *et al*. Antimicrobial misuse in patients with positive blood cultures. *Am J Med* 1989; **87**: 253–9.
39. Siegel JD, Rhinehart E, Jackson M, Chiarello L, the Healthcare Infection Control Practices Advisory Committee. *Management of Multidrug-resistant Organisms in Healthcare Settings*, 2006. http://www.cdc.gov/hicpac/mdro/mdro_0.html (accessed July 2011).
40. D'Agata EM. Methodologic issues of case-control studies: a review of established and newly recognized limitations. *Infect Control Hosp Epidemiol* 2005; **26**: 338–41.
41. Hyle EP, Bilker WB, Gasink LB, Lautenbach E. Impact of different methods for describing the extent of prior antibiotic exposure on the association between antibiotic use and antibiotic-resistant infection. *Infect Control Hosp Epidemiol* 2007; **28**: 647–54.
42. Royston P, Altman D, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006; **25**: 127–41.
43. MacAdam H, Zaoutis TE, Gasink LB, Bilker WB, Lautenbach E. Investigating the association between antibiotic use and antibiotic resistance: Impact of different methods of categorizing prior antibiotic use. *Int J Antimicrob Agents* 2006; **28**: 325–32.
44. Gasink LB, Zaoutis TE, Bilker WB, Lautenbach E. The categorization of prior antibiotic use: Impact on the identification of risk factors for drug resistance in case control studies. *Am J Infect Control* 2007; **35**: 638–42.
45. Lee I, Morales KH, Zaoutis TE, Fishman NO, Nachamkin I, Lautenbach E. Clinical and economic outcomes of decreased fluconazole susceptibility in patients with *Candida glabrata* bloodstream infections. *Am J Infect Control* 2010; **38**: 740–5.
46. Merrer J, Santoli F, Appere de Vecchi C, Tran B, De Jonghe B, Outin H. "Colonization pressure" and risk of acquisition of methicillin-resistant Staphylococcus aureus in a medical intensive care unit. *Infect Control Hosp Epidemiol* 2000; **21**: 718–23.
47. Harris AD, Lautenbach E, Perencevich E. A systematic review of quasi-experimental study designs in the fields of infection control and antibiotic resistance. *Clin Infect Dis* 2005; **41**: 77–82.
48. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company, 2002.
49. Cook TD, Campbell DT. *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally Publishing, 1979.
50. Harris AD, Bradham DD, Baumgarten M, Zuckerman IH, Fink JC, Perencevich EN. The use and interpretation of quasi-experimental studies in infectious diseases. *Clin Infect Dis* 2004; **38**: 1586–91.
51. Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. *BMJ* 2003; **326**: 1083–4.
52. Harris AD, Carmeli Y, Samore MH, Kaye KS, Perencevich E. Impact of severity of illness bias and control group misclassification bias in case-control

studies of antimicrobial-resistant organisms. *Infect Control Hosp Epidemiol* 2005; **26**: 342–5.

53. Kaye KS, Harris AD, Samore M, Carmeli Y. The case-case-control study design: addressing the limitations of risk factor studies for antimicrobial resistance. *Infect Control Hosp Epidemiol* 2005; **26**: 346–51.
54. Harris AD, Karchmer TB, Carmeli Y, Samore MH. Methodological principles of case-control studies that analyzed risk factors for antibiotic resistance: a systematic review. *Clin Infect Dis* 2001; **32**: 1055–61.
55. Harris AD, Samore MH, Lipsitch M, Kaye KS, Perencevich E, Carmeli Y. Control-group selection importance in studies of antimicrobial resistance: examples applied to *Pseudomonas aeruginosa*, *Enterococci*, and *Escherichia coli*. *Clin Infect Dis* 2002; **34**: 1558–63.
56. Carmeli Y, Samore MH, Huskins C. The association between antecedent vancomycin treatment and hospital-acquired vancomycin-resistant enterococci. *Arch Intern Med* 1999; **159**: 2461–8.
57. Gasink LB, Blumberg EA, Localio AR, Desai SS, Israni AK, Lautenbach E. Hepatitis C virus seropositivity and survival in heart transplant recipients. *JAMA* 2006; **296**: 1843–50.
58. 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.
59. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. *Int J Technol Assess Health Care* 2003; **19**: 613–23.
60. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther* 2002; **27**: 299–309.
61. Matowe LK LC, Crivera C, Korth-Bradley JM. Interrupted time series analysis in clinical research. *Ann Pharmacother* 2003; **37**: 1110–6.
62. Stone SP, Cooper BS, Kibbler CC, Cookson BD, Roberts JA, Medley GF, *et al*. The ORION statement: guidelines for transparent reporting of outbreak reports and intervention studies of nosocomial infection. *Lancet Infect Dis* 2007; **7**: 282–8.
63. Kaye KS, Harris AD, Gold H, Carmeli Y. Risk factors for recovery of ampicillin-sulbactam-resistant *Escherichia coli* in hospitalized patients. *Antimicrob Agents Chemother* 2000; **44**: 1004–9.
64. Harris AD, Smith D, Johnson JA, Bradham DD, Roghmann MC. Risk factors for imipenem-resistant *Pseudomonas aeruginosa* among hospitalized patients. *Clin Infect Dis* 2002; **34**: 340–5.
65. Harris AD, Perencevich E, Roghmann MC, Morris G, Kaye KS, Johnson JA. Risk factors for piperacillin-tazobactam-resistant Pseudomonas aeruginosa among hospitalized patients. *Antimicrob Agents Chemother* 2002; **46**: 854–8.

CHAPTER 45
The Pharmacoepidemiology of Medication Errors

Hanna M. Seidling[1] *and David W. Bates*[2]

[1] Department of Clinical Pharmacology and Pharmacoepidemiology, Cooperation Unit Clinical Pharmacy, University of Heidelberg, Germany

[2] Department of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, USA

Introduction

Background

Medications are 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] while in the hospital multiple medication orders tend to be written for each patient daily. Medication errors are frequent, but fortunately only a small proportion result in harm.[2] However, given the prevalence of prescription medication use, preventable adverse drug events are one of the most frequent causes of preventable iatrogenic injuries. The Institute of Medicine report, "To Err is Human," suggested at least 44 000–98 000 deaths per year occur in the US from iatrogenic injury.[3] One study estimated that about 7000 deaths are attributed to medication errors exclusively[4] and about 1 million injuries might result from drug 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 (i.e., knowledge-based mistakes or rule-based mistakes) or errors in executing correctly planned actions (i.e., action-based slips or memory-based lapses).[6] 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, and these errors may be intercepted before they reach the patient, or reach the patient without consequence. However, generally, about one in ten 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 45.1). 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 been defined as harm that is caused by a drug while appropri-

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

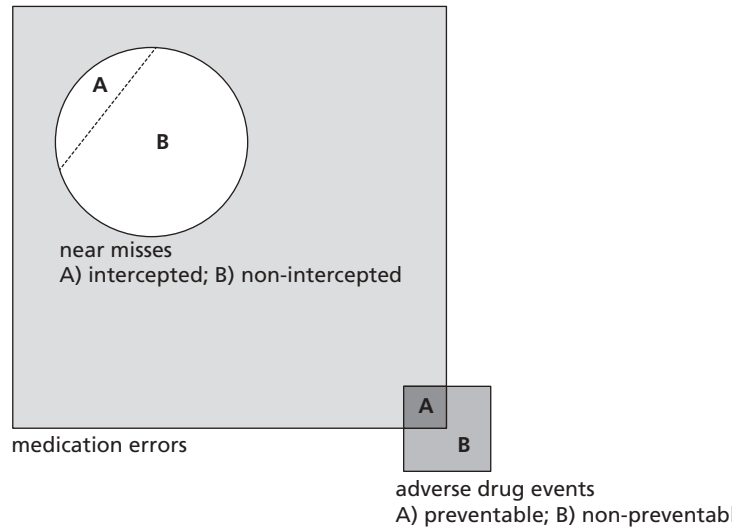


Figure 45.1 Relationship between medication errors and adverse drug events. About 1 in 10 medication errors is likely to result in patient harm,[7] whereas about 25% of adverse drug events can be allocated to a medication error.[2] Near misses, both intercepted and non-intercepted, comprise those medication errors with potential for patient harm without resulting in actual harm.

ately used.[9] (See also Chapter 1 for alternative definitions.)

Detection of medication errors

These approaches include manual or automatic screening of claims data, administrative databases, medical records, electronic health records, incident reports mostly by providers in hospitals, patient monitoring, direct observation often by pharmacists, and spontaneous (self-reporting) approaches. All of these 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 that 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, 10, 30, 33, and 46 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 data sets, but is generally done retrospectively. The quality of the available information, however, varies between different data sources, which restricts opportunities to comprehensively detect medication errors—and this applies more to some types of medication errors than others. 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. Weaknesses include that it cannot be determined with certainty whether or not the patient actually consumed the medication, and, if not linked to other information sources, clinical detail is often minimal (e.g., information on weight or renal function might be missing), making it hard to answer questions that relate to a patient's clinical condition (see Part IIIb for additional discussion of automated data systems). In particular, 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 usually will not be 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. The main problems with chart review are that it is time-consuming and expensive, with the average chart review costing approximately $20. When electronic health records are in place, the manual screening of paper-based information can be replaced by semiautomated approaches. The level of standardization and the extent to which clinical information is stored by using controlled vocabulary determines the feasibility and effectiveness of automated, algorithm-based data analyses.[11]

When electronic health records include electronic prescribing applications with clinical decision support (i.e., CPOE- computerized physician order entry), 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 actually resulted in patient harm. Screening of incident reports always underestimates the incidence of errors by a large degree (because of under-reporting of events), but is relatively inexpensive, because data are collected as a byproduct of routine care delivery. Patient monitoring for adverse drug events has been successful, and can identify more adverse drug events than chart review.[13]

Spontaneous (self-reported) reporting of medication errors is comparatively easy to be set in place and to maintain, in both inpatient and outpatient settings. However, both ADEs and medication errors are substantially under-reported (see Chapter 10). Indeed, the major barrier for reporting medication errors is the perception by staff that reporting might be associated with disciplinary actions, even if the hospital pursues a non-punitive policy.[14] Thus, spontaneous reporting is only useful for obtaining samples of errors and cannot be used to assess the underlying rate of medication errors in a sample.[15]

Direct observation is primarily done during research studies at inpatient sites 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,[16] and it is particularly useful at stages that are not sensitive to other detection methods (e.g., drug preparation or drug administration).[17]

Methodologic problems to be addressed by pharmacoepidemiologic research

Pitfalls in the detection of medication errors

The reliable and comprehensive detection of medication errors has a number of methodologic problems, 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 or not there was harm or potential for harm, and the decision about whether or not to include errors that are intercepted even before reaching the patient.

Identification of medication errors also remains challenging, and general standards are lacking. For instance, the detection of wrong timing errors requires the definition of a threshold value above which the medication is considered to be delayed. In the inpatient setting, this threshold value might be sometimes 2 or 4 hours. Moreover, sometimes patients are away from their inpatient rooms (e.g., getting diagnostic tests), in which case decisions

need to be made about whether or not to keep the threshold values the same.

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 knowledge base. However, the content of such knowledge bases varies widely, both in terms of included drug pairs as well as specific information linked to a drug pair (e.g., severity of the drug–drug interaction).[18] Especially in the outpatient setting, comprehensive and reliable data on the actual patient's medication list may be missing; prescribing and dispensing data are seldom jointly available, and determining patient adherence to whatever drug they take at home is even more difficult. Even patient surveys may not give adequate information—while patients might be nonadherent to some prescribed drugs, they might also be taking over-the-counter drugs with potential for drug–drug interactions (e.g., St John's wort) which they do not report.[19]

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 or not the patient is actually pregnant at the time of the exposure. In retrospective analyses, identification of the date of birth, and backward calculation under the assumption of a term pregnancy might be feasible, though this can be complicated since such information is not readily stored in one location. Information on whether a woman is pregnant or not at the time of prescribing is challenging to obtain and even most information systems do not have good approaches for tracking this.

Another important piece of clinical information, especially in pediatrics (though also for chemotherapy and some other situations), is the patient's weight. Most pediatric medications are dosed on the basis of weight. Standardized documentation of this information can be challenging to obtain, hindering not only analyses of pediatric dosing but also actual dosing by pediatricians. Obtaining an accurate weight is also essential for oncology, as it is in dosing of certain intravenous drugs, especially in obese patients.

Finally, information on patients' allergies is only infrequently and inconsistently available.[20,21] It is important that allergies be differentiated from sensitivities or intolerances 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 allergy list in an electronic format for each patient, rather than multiple disparate lists.

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 incidence rates determined in different studies has substantial limitations, most prominently the fact that different detection approaches and also different numerators and denominators may be used. Thus, medication error rates from different studies can be difficult to compare unless the same, or similar, detection approaches were used. Other factors to consider are the setting studied and the patient population.

Medication error rates by setting

The vast majority of the early 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, as well as at times due to 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 non-university hospitals. A recent review 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 non-university hospitals.[7] Studies of ADE rates in hospitals have found incidence rates ranging from 2 to 15 per 100 admissions.[5,22,23]

In intensive care units (ICUs), the rates of medication errors appear higher than on general care

units; many more medications are used and they have higher levels of toxicity. Beyond the frequency, the nature and causes of medication errors are different and the risk that a medication error actually will result in patient harm is also higher in the ICU than on general inpatient wards,[24] with 7.4% of patients experiencing an ADE resulting from a medication error.[25]

In nursing homes, and especially in the ambulatory setting, assessment of medication error incidence is challenging because individual steps in the medication process are rarely jointly documented, and there are often substantial time lags in between them. Sometimes estimation of frequency of medication errors has relied on spontaneous self-reporting systems (resulting in dramatic underestimates of frequency)[26] or documentation of ADEs in charts, which misses both many ADEs and also nearly all medication errors.

In the ambulatory setting, patients live in their homes and take their medications independently, which makes detection of medication errors and ADE challenging. However, the incidence of ADEs can be estimated by direct patient surveys, and the most severe ADEs can be assessed from frequency of hospital admissions resulting from ADEs. ADE rates range from 25% of patients (as self-reported in a survey)[13] to 5% (of hospital admissions).[27] However, comprehensive data on the incidence rates of medication errors in the ambulatory setting are lacking. As for prescribing errors exclusively, rates are reported as 7.6% of all prescription orders.[28] Medication error rates stratified for different specializations or dentists have not been studied in detail.[29]

Another issue is what happens at the interfaces of care, although it is clear that transitions may be especially risky. Only a few studies have been conducted to assess the incidence and nature of medication errors at the interface between primary and tertiary care,[30] and, especially in the elderly population, the incidence of problems with the drug prescription regimen are frequent after discharge (in about one-third of elderly, discharged patients) and contribute to higher re-hospitalization rates.[31]

Medication error rates in 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.[2] Seven in 100 medication errors had significant potential for harm, and 1 in 100 actually resulted in an injury.[2]

Medication error rates in pediatric patients are estimated to be as high as 5–27% of all medication orders,[32] most of those studies having been performed in the inpatient setting. In neonatal intensive care units, error rates have been reported to be in similar ranges.[33] 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).[34]

ADE rates have also been reported for the elderly; as many as 35% of elderly outpatients per year may experience an ADE,[35] and about 30% of hospital admission are ADE-related in the elderly.[36] In elderly patients, many medication error studies have been done focusing on the prescription of inappropriate drugs, especially using the Beers criteria (i.e., a list of drugs specified through consensus of experts that should be avoided in elderly patients in general or under consideration of specific cofactors including co-morbidity or dosage),[37] although the utility of these criteria has been challenged.[38]

Medication error rates by detection method

The incidence of medication errors may vary as much as a 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.[39] 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 training grade and profession of the person who conducts the detection.[39]

In general, medication error incidence rates are underestimated by as much as 10 000-fold if spontaneous reporting methods are applied.[40] In order to promote spontaneous reporting, non-punishment policies as well as anonymous reporting have been established. Moreover, it is especially crucial to invite those professions that 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 all their medications from one single pharmacy, many medication errors become evident in the pharmacy and not during a doctor's consultation. Thus, pharmacists should be invited to report medication errors to identify a greater number of such errors.[41]

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 instance, the susceptibility to suffer an ADE is higher in geriatric wards as well as ICU patients compared to general care units (12 vs. 6%).[24] 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.[42] 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] Non-fatal ADEs might prolong the hospital stay or increase the risk of re-hospitalization. In another study, 13% of patients experienced an ADE after discharge, and of these 24% were preventable and 38% ameliorable.[43] In addition, ADEs occurring in the outpatient setting can contribute to hospital admissions, with 4.5 preventable ADEs per 1000 person-months.[44]

Gaps in clinical care promoting medication errors

General risk factors

The search for risk factors for medication errors has been challenging, as some appear to occur relatively randomly in the medication process, and robust systems need to detect and prevent even errors occurring randomly.[45] However, some situations that are associated with a higher risk of error can be identified. In particular, it can be helpful to determine: (i) at what stage of the treatment process medication errors are occurring, (ii) 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, (iii) what the patient's characteristics are, including age, comorbidities, and other medications they are taking, and (iv) what the clinical setting is.

Well-defined risk factors include renal dysfunction and old age, which is associated with renal dysfunction and prescribed drug classes. Setting is also important, with ICU patients having an especially high risk, because they are more seriously ill but also because they are exposed to large numbers of medications. Other risk factors can make persons involved in the treatment process more susceptible to commit an error, for example, knowledge about the clinical condition and drug, and workload and stress level, which when increased are associated with higher rates of slips and lapses.

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.[46] However, in all settings, wrong dose is the most frequent type of medication error, especially over-dosage. Dosage errors may occur at the stage of administration (accidental intake of two tablets), the stage of manufacturing or dispensing (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 co-medication, and genetic polymorphisms. Moreover, required dosages will depend on age-related pharmacodynamic changes, vary between disease conditions that are intended to be treated, and might 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 both in under- dosage, or more often in over-dosage.

Drug classes strongly associated with medication errors

Any drug or drug formulation can be associated with a medication error. However, there are some active ingredients that are particularly frequently connected with medication errors. Predisposing factors include: (i) a sophisticated way of prescribing (e.g., complex dosage adjustments), administration (e.g., usage of administration devices), or monitoring (e.g., therapeutic drug monitoring); (ii) a substantial dose-dependent toxicity, which increases the likelihood that a medication error will indeed result in patient harm; and (iii) a prescription frequency that is high enough that the error will occur during study periods but low enough that drug handling remains challenging.

The drug class with the highest prescription frequency is cardiovascular drugs. Indeed, in many medication error studies, cardiovascular drugs have often been reported to be involved in medication errors and ADEs.[47] Moreover, the prescription of antibiotics also often resulted in ADEs, most often because known allergies were ignored.[47] Medication errors with fatal outcomes, however, are often associated with drugs that are less frequently used but complicated in their mode of administration. For instance, accidental intrathecal injection of vincristine has caused several dozens of deaths[48] and even though many measurements for error prevention have been undertaken, the error still occurs.[49] 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 a three- to fourfold higher maximum recommended doses for the liposomal preparation. Erroneous administration of aqueous amphotericin B solution in dosages appropriate only for the liposomal preparation have resulted in a number of cases in renal toxicity and death.[50]

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 in order to maintain effectiveness and prevent ADRs such as bleeding. In one inpatient study,[51] 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.[52]

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.[53]

Specific gaps by setting

Inpatient setting

In adult inpatients, wrong dosage resulting from knowledge-based errors is the most frequent medication error. Patients suffer from multiple comorbidities, and some may require dosage adjustment of which the prescribing physician is unaware at the time of prescribing. In pediatric inpatients, wrong dosage often results from calculation errors, including 10-fold errors.[54] 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 countries, the majority of such potential medication errors are intercepted by hospital pharmacists while processing the order.

Outpatient setting

In the outpatient setting, many medication errors happen at the stage of drug monitoring (e.g.,

neglecting a required checkup 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. In the outpatient setting in the elderly, the number of physicians seen by a patient was found as independent risk factor for an ADE.[55] 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. Thus, prescription of interacting drugs is frequent and drug–drug interactions contribute to 6% of ADE-related hospital admissions.[27] Compared to the inpatient setting, prescription errors are less likely to be intercepted and, moreover, the patient plays a more active role in their medical treatment. Hence, responsibility for appropriate drug administration lies with the patient. Two major factors might impede appropriate drug administration: (i) patient non-adherence to prescribed drugs (see also Chapter 42), and (ii) 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.[56] Additionally, inappropriate splitting of tablets was found to be the source of some medication errors.[57]

Intensive care unit

In the ICU, critically ill patients are characterized by rapidly changing clinical conditions, they receive close and intensive monitoring, and require rapid adaptations of their drug therapies. Due to the large number of necessary medications, the frequency of drug–drug interactions is particularly high, with about two-thirds of patients having at least one drug–drug interaction and 44% suffering from a drug–drug interaction-related ADR in one study.[58] 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.[59]

Long-term care

In the long-term care setting, relatively few data are available.[60] However, medication errors appear to be concentrated in a few different drug classes, most often involving drugs affecting the central nervous system or analgesics.[51] Pharmacotherapy in the elderly occurs in a patient population that is in general multimorbid, polymedicated, and has physiological changes requiring complex dosage adjustment. Therefore, prescribing errors involving inappropriate drug choice as well as inappropriate dosages are frequent.[61]

In conclusion, a multitude of different combination of risk factors is possible and each set of risk factors will require a distinct prevention strategy to effectively prevent medication errors.

Currently available solutions

Options for prevention of medication errors

Examples of prevention strategies

Obviously, the best prevention strategies for medication errors will depend on the setting, and in particular the nature of the medication errors involved. Slips and lapses in executing correctly planned actions can be addressed by workflow changes, including skills training and monitoring (dual control systems, actions acknowledged by a second pair of eyes, checklists).[62] In contrast, mistakes may be prevented by providing relevant knowledge at the time it is required. Approaches could 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 (computerized physician order entry or CPOE) with integrated clinical decision support systems (CDSS) assumes a key role in medication error prevention.[63] Implementation of such systems

might eliminate several types of errors, such as transcribing errors,[64] and reduce others, although their impact on ADEs has been more variable,[65] partly due to the fact that most of the studies that have been performed have been underpowered. Electronic solutions have also been developed in order to safeguard drug dispensing or administration. Most prominent among these are bar-coding systems, which are aimed at reducing administration of medications to the wrong patient.[66] Electronic medication administration records can be used to electronically monitor drug administration and effectively reduce errors of omission.[67]

Options regarding outcome assessment

The outcome of prevention strategies is often reported as changes in the frequency of medication errors. However, such information has only limited application as a predictor for health-related outcomes. Indeed, in studies assessing both medication errors and patient outcome, a reduction in medication errors would not necessarily be accompanied by an improvement in patient outcome. For example, a computer-assisted disease management system might enhance the number of guideline-conformed screenings; however, the disease severity would not be ameliorated.[68] The assessment of patient outcome, 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. Moreover, most prevention strategies are evaluated in a before versus after implementation setting, and are only rarely evaluated in randomized trials. Therefore, neglect 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 increasing the mortality rates of pediatric inpatients (odds ratio 3.28; 95% confidence interval: 1.94–5.55).[69] 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,[70,71] 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.[72]

The future

Conclusion and prospects

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 make it hard to narrowly define their incidence and severity. The frequency, best detection approaches, and prevention methods vary by setting and patient population. In order to allow comparison of single study results, the definition of consistent numerator and denominator is especially important. Especially in large-scale studies using only administrative data, information relevant to the reliable identification of medication errors is often not comprehensively available. Key steps are that the health research community agrees on and advances: (i) consistent use of definitions and classifications of medication errors, and (ii) attempts to merge large medication databases with electronic data on patient's clinical information. 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. 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. Kohn LT, Corrigan JM, Donaldson MS. Institute of Medicine. *To Err is Human. Building a Safer Health System*. Washington, D.C.: National Academy Press, 1999.
4. Starfield B. Is the US health really the best in the world. *J Am Med Assoc* 2000; **284**: 483–5.
5. Bates DW, Cullen D, Laird N, Petersen LA, Small S, Servi D, *et al*. Incidence of adverse drug events and potential adverse drug events: implications for prevention. *J Am Med Assoc* 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. *Proc AMIA Symp* 2002: 106–10.
12. Seidling HM, Schmitt SP, Bruckner T, Kaltschmidt J, Pruszydlo MG, Senger C, *et al*. Patient-specific electronic decision support reduces prescription of excessive doses. *Qual Saf Health Care* 2010; **19**: e15.
13. Gandhi TK, Weingart SN, Borus J, Seger AC, Peterson J, Burdick E, *et al*. Adverse drug events in ambulatory care. *N Engl J Med* 2003; **348**: 1556–64.
14. Potylycki MJ, Kimmel SR, Ritter M, Capuano T, Gross L, Riegel-Gross K, *et al*. Nonputitive medication error reporting: 3-year findings from one hospital's primum non nocere initiative. *J Nurs Adm* 2006; **36**: 370–6.
15. 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.
16. 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.
17. Dean B, Barber N. Validity and reliability of observational methods for studying medication administration errors. *Am J Health Syst Pharm* 2001; **58**: 54–9.
18. Metzger J, Welebob E, Bates DW, Lipsitz S, Classen DC. Mixed results in the safety performance of computerized physician order entry. *Health Aff* 2010; **29**: 655–63.
19. Rieger K, Scholer A, Arnet I, Peters FT, Maurer HH, Walter-Sack I, *et al*. High prevalence of unknown co-medication in hospitalised patients. *Eur J Clin Pharmacol* 2004; **60**: 363–8.
20. Kuperman GJ, Gandhi TK, Bates DW. Effective drug-allergy checking: methodological and operational issues. *J Biomed Inform* 2003; **36**: 70–9.
21. 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.
22. 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. *J Am Med Assoc* 1997; **277**: 301–6.
23. Hug BL, Witkowski DJ, Sox CM, Keohane CA, Seger DL, Yoon C, *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.
24. 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.
25. Benkirane RR, Abouqal R, Haimeur CC, S Ech Cherif El Kettani SS, Azzouzi AA, M'daghri Alaoui AA, *et al*. Incidence of adverse drug events and medication errors in intensice care units: a prospective multicenter study. *J Patient Saf* 2009; **5**: 16–22.
26. 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.
27. McDonnell PJ, Jacobs MR. Hospital admissions resulting from preventable adverse drug reactions. *Ann Pharmacother* 2002; **36**: 1331–6.
28. Gandhi TK, Weingart SN, Seger AC, Borus J, Burdick E, Poon EG, *et al*. Outpatient prescribing errors and the impact of computerized prescribing. *J Gen Intern Med* 2005; **20**: 837–41.

29. Medonca JM, Lyra DP Jr, Rabelo JS, Siqueira JS, Balisa-Rocha BJ, Gimenes FR, *et al*. Analysis and detection of dental prescribing errors at primary health care units in Brazil. *Pharm World Sci* 2010; **32**: 30–5.
30. 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.
31. 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.
32. 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.
33. Chedoe I, Molendijk HA, Dittrich ST, Jansman FG, Harting JW, Brouwers JR, *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.
34. Walsh KE, Dodd KS, Seetharaman K, Roblin DW, Herrinton LJ, Von Worley A, *et al*. Medication errors among adults and children with cancer in the outpatient setting. *J Cin Oncol* 2009; **27**: 891–6.
35. Hanlon JT, Schmader KE, Koronkowski MJ, Weinberger M, Landsman PB, Samsa GP, *et al*. Adverse drug events in high risk older outpatients. *J Am Geriatr Soc* 1997; **45**: 945–8.
36. Onder G, Pedone C, Landi F, Cesari M, Della Vedova C, Bernabei R, *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.
37. 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.
38. 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.
39. 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.
40. 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.
41. 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.
42. Kaushal R, Bates DW, Landrigan C, McKenna KJ, Clapp MD, Federico F, *et al*. Medication errors and adverse drug events in pediatric inpatients. *J Am Med Assoc* 2001; **285**: 2114–20.
43. 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.
44. 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.
45. Bates DW, Miller EB, Cullen DJ, Burdick L, Williams L, Laird N, *et al*. Patient risk factors for adverse drug events in hospitalized patients. ADE Prevention Study Group. *Arch Intern Med* 1999; **159**: 2553–60.
46. 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.
47. 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.
48. 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**: 816–9.
49. Noble DJ, Donaldson LJ. The quest to eliminate intrathecal vincristine errors: a 40-year journey. *Qual Saf Health Care* 2010; **19**: 323–6.
50. 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.
51. 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.
52. Hansen RA, Grenne SB, Williams CE, Blalock SJ, Crook KD, Akers R, *et al*. Types of medication errors

in North Carolina nursing homes: a target for quality improvement. *Am J Geriatr Pharmacother* 2006; **4**: 52–61.

53. 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.
54. Lesar TS. Tenfold medication dose prescribing errors. *Ann Pharmacother* 2002; **36**: 1833–39.
55. 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.
56. 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.
57. Quinzler R, Szecsenyi J, Haefeli WE. Tablet splitting: patients and physicians need better support. *Eur J Clin Pharmacol* 2007; **63**: 1203–4.
58. Bertsche T, Pfaff J, Schiller P, Kaltschmidt J, Pruszydlo MG, Stremmel W, *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.
59. 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.
60. 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.
61. Fialova D, Onder G. Medication error in elderly people: contributing factors and future perspectives. *Br J Clin Pharmacol* 2009; **67**: 641–5.
62. Conroy S, Sweis D, Planner C, Yeung V, Collier J, Haines L, *et al*. Intervention to reduce dosing errors in children: a systematic review of the literature. *Drug Saf* 2007; **30**: 1111–25.
63. Anton C, Nightingale PG, Adu D, Lipkin G, Ferner RE. Improving prescribing using a rule based prescribing system. *Qual Saf Health Care* 2004; **13**: 186–90.
64. Bates DW, Leape LL, Cullen DJ, Laird N, Petersen LA, Teich JM, *et al*. Effect of computerized physician order entry an a team intervention on prevention of serious medication errors. *J Am Med Inform Assoc* 1998; **280**: 1311–6.
65. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano ML, Devereaux PH, Beyene J, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *J Am Med Assoc* 2005; **293**: 1223–38.
66. Anderson S, Wittwer W. Using bar-code point-of-care technology for patient safety. *J Health Qual* 2004; **26**: 5–11.
67. 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.
68. Meigs JB, Cagliero E, Dubey A, Murphy-Sheehy P, Gildesgame C, Chueh H, *et al*. A controlled trial of Web-based diabetes disease management: the MGH Diabetes Primary Care Improvement Project. *Diabetes Care* 2003; **26**: 750–7.
69. Han YY, Carcillo JA, Venkataraman ST, Clark RS, Watson RS, Nguyen TC, *et al*. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 2005; **116**: 1506–12.
70. 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.
71. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, *et al*. Role of computerized physician order entry systems in facilitating medication errors. *J Am Med Assoc* 2005; **293**: 1197–203.
72. 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.

CHAPTER 46
Sequential Statistical Methods for Prospective Postmarketing Safety Surveillance

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

Introduction

Near real-time postmarketing drug and vaccine safety surveillance systems are increasingly being developed and used to quickly detect potential safety problems. Many of these systems use automated weekly or monthly data updates from electronic medical health records or health insurance claims, which contain information about both drug/vaccine exposure and potential adverse event outcomes from a well-defined population. When statistical analyses are repeatedly conducted on the same, slightly expanded data, sequential statistical methods are needed to adjust for the multiple testing inherent in the many looks at the data. This ensures that the correct alpha level (type 1 error, see Chapter 4) is maintained throughout the surveillance period, so that the probability of a false positive is at the desired level.

In this chapter, we describe different sequential analysis methods that have been used for prospective postmarketing drug safety surveillance; we show how they have been utilized for different vaccines and adverse events and with different types of comparison groups; and we discuss different sequential design options that are important to consider when designing a prospective, near real-time safety surveillance system.

Clinical problems to be addressed by pharmacoepidemiologic research

When a new medical product enters the market, whether it is a drug, vaccine, or device, there are always questions about its safety (see Chapter 1). Postmarketing safety surveillance is important in order to detect serious adverse events that are not detected in premarketing clinical trials, either because they are too rare or because they only affect a subpopulation that was excluded from the trial, such as pregnant women (see Chapter 28). Postmarketing surveillance has traditionally been based on spontaneous adverse event reporting systems[1] (see Chapter 10), but huge observational electronic health data sets from health insurance plans are increasingly available and used as well[2–4] (see Chapters 11–18). For example, as part of the new, US, Congressionally mandated Sentinel Initiative (see Chapter 30), a population of over 100 million will be gathered, specifically to screen for adverse effects. When such surveillance is conducted repeatedly over time, with multiple analyses of the same data set as the data accrues, sequential statistical methods should be used.

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

Methodologic problems to be solved by pharmacoepidemiologic research

If there is a safety problem with a medical product, it is important to know about it as soon as possible. Data from spontaneous reporting systems are often looked at repeatedly over time through the repeated calculation of, for example, proportional reporting ratios,[5] reporting odds ratios,[6] Bayesian confidence propagation neural networks,[7] or empirical Bayes gamma Poisson shrinkers[8,9] (see Chapter 10). Electronic health data can also be frequently accessed to continuously monitor the safety of newly approved drugs or vaccines[10,11] (see Chapter 30). It is perfectly valid to monitor point estimates of the above measures continuously, but in order to avoid an excess number of false alarms when doing hypothesis testing or calculating confidence intervals, it is important to adjust for the multiple comparisons inherent in the frequent analyses performed as additional data accrue.

Sequential statistical methods allow a statistical signal to be generated as soon as there is sufficient evidence to reject the null hypothesis that there is no excess risk of the adverse event, while also ensuring that the probability of falsely rejecting the null hypothesis at any time during the surveillance is controlled at the desired nominal significance level.[12,13] They can be broadly categorized as continuous sequential methods, which allow testing to be performed continuously over time with as many analyses of the data as the investigator desires, and group sequential methods, which involve analyzing data at regular or irregular discrete time intervals after a group of subjects enter the study.

Sequential analysis versus control chart methods

There are two different types of statistical methods that are commonly used for repeated analyses on accruing data: sequential analysis methods, such as the sequential probability ratio test, and control chart methods, such as the cumulative sum chart (CUSUM).[14] Both types are used for monitoring something over time, but their purposes are very different,[15] and it is important to know when to use which one. Sequential analysis methods are designed to quickly detect a problem that has always been there from the beginning of the analyses, such as an adverse event caused by the inherent properties of the drug. Chart-based methods, on the other hand, are designed to monitor a process for a sudden shift or change that occurs at some unknown time in the future, with the goal of detecting it as soon as possible after the problem occurs. They do so by comparing the number of events observed during recent time with a baseline event rate from earlier times. Chart-based methods are commonly used in an industrial setting to quickly detect a suddenly malfunctioning manufacturing process, and they can be used to detect adverse events that are due to a suddenly emerging manufacturing problem of a normally safe medication. In this chapter, though, we focus exclusively on the detection of adverse events due to inherent problems with correctly manufactured drugs, so, from here on, we only consider sequential analysis methods.

Example of a simple sequential surveillance set-up

We begin by presenting a simple Poisson-based data model with observed and expected counts of a predefined vaccine adverse event. For each person vaccinated, determine whether an adverse event occurred during the D days following vaccination. Let v_t be the number of persons vaccinated during the time period $[0,t]$, and let c_t be the number of these persons that had the adverse event within D days of vaccination. Note that time t is defined relative to the time of vaccination rather than the time of the adverse event. Hence, we actually do not know the value of c_t until time $t+D$.

Under the null hypothesis that there is no excess risk of the adverse event due to the vaccine, there is still some probability q that a vaccinated person will be diagnosed with the adverse event within D days of the vaccination, just by chance. The value of q, which for the time being is assumed to be known, can be calculated in a wide variety of ways depending on the specific surveillance setting. For now, we can think of it as being based on a large number of historical controls or from incidence

rates published in the literature. If the adverse event is rare, q is very small, and under the null hypothesis c_t follows a Poisson distribution with mean $\mu_t = q*v_t$ (H_0:$RR = E[c_t]/\mu_t = 1$). To adjust for covariates such as age, a different probability q_i is used for each age group, and $\mu_t = \Sigma_i\, q_i*v_{ti}$, where v_{ti} is the number of vaccinated persons in age group i during the time period $[0,t]$. Note that $c_0 = \mu_0 = 0$, and that both the expected and observed number of adverse events increases with t as more individuals are vaccinated.

The data are monitored continuously over time as additional data are collected. If, at some point in time, the observed number of adverse events is considerable larger than the expected number, there is evidence against the null hypothesis. Different sequential analysis methods use different functions of t, μ_t, and c_t to define a test statistic, and when this test statistic exceeds a predetermined critical value, a statistical signal is generated rejecting the null hypothesis. The critical value is chosen so that if the null hypothesis of no excess risk due to the vaccine is true, then the probability of rejecting the null hypothesis at any time during the surveillance is $\alpha = 0.05$, or some other chosen value. Hence, we can maintain the desired probability of generating a false-positive signal at any time during the surveillance, adjusting for the multiple testing inherent in the many repeated analyses performed as the data accrues.

Currently available solutions

Continuous sequential analysis methods

Wald's sequential probability ratio test (SPRT)

Continuous sequential methods were introduced by Wald in the 1940s, who proposed the sequential probability ratio test (SPRT), which can be used for a wide variety of probability models.[12,13] For the setting described above, the alternative hypothesis is that c_t follows a Poisson distribution with the mean $r*\mu_t$ for some fixed and predetermined value of r, such as $r = 3$ (H_A: $RR = 3$). Since the alternative hypothesis only contains a single possible value for the parameter r, it is a singular alternative. The test statistic is the likelihood ratio, which for the Poisson distribution is defined as

$$LR_t = \frac{P(c_t \mid HA)}{P(c_t \mid H0)} = \frac{e^{-r\mu_t}(r\mu_t)^{c_t}/c_t!}{e^{-\mu_t}\mu_t{}^{c_t}/c_t!} = e^{(1-r)\mu_t} r^{c_t}$$

In words, the likelihood ratio test statistics calculates the probability of observing what was observed both under the alternative hypothesis, where $RR = r$, and under the null hypothesis, where $RR = 1$. If the probability of the former is higher than the latter, there is more evidence for the alternative hypothesis. With a bigger difference, the likelihood ratio is larger, and there is more evidence in favor of the alternative hypothesis over the null. Hence, the null hypothesis is rejected when the likelihood ratio is large and the alternative hypothesis is rejected when it is small.

Because it is numerically easier to calculate, the log likelihood ratio

$$LLR_t = (1-r)\mu_t + c_t \ln(r)$$

is typically used instead of the likelihood ratio, and a since the logarithm is a monotone function that maintains the ordering of all values, the two are equivalent test statistics.

The log likelihood ratio based test statistic is continuously monitored for all values of $t > 0$, until either $LLR_t \geq ln((1-\beta)/\alpha)$, in which case the null hypothesis is rejected, or until $LLR_t \leq ln(\beta/(1-\alpha))$, in which case the alternative hypothesis is rejected. With this stopping boundary, the null hypothesis will be falsely rejected with probability α when it is true (type 1 error) while the alternative hypothesis will be falsely rejected with probability β when it is true (type 2 error), although it should be noted that these are approximate results.[12,13] Note that $LLR_0 = 0$.

The classical SPRT is not well suited for pharmacoepidemiologic surveillance, due to the singular (e.g., H_A:$RR = 3$) rather than composite (H_A:$RR > 1$) nature of the alternative hypothesis.[16,17] If the alternative hypothesis specifies a high relative risk, such as $RR = 3$, but the true relative risk is low, such as $RR = 1.5$, the classical SPRT may accept rather than reject the null hypothesis since the null is closer to

the truth than the alternative. On the other hand, if the alternative specifies a low relative risk, such as $RR = 1.5$, while the true relative risk is high, such as $RR = 3$, the method may take a long time to reject the null since the null and alternative hypotheses are almost equally distant from the truth.[17]

Abt's sequential test

Abt[16] recognized the above problem with the classical SPRT and proposed an alternative. For some values of a and b specified by the user, with $0 < a < 1$ and $0 < b < 1$, define

$$R(a,b,t) = RR : \frac{\ln((1-b)/a) + \mu_t(RR-1)}{\ln(RR)}$$
$$= \min_{RR>1} \frac{\ln((1-b)/a) + \mu_t(RR-1)}{\ln(RR)}$$

This means that, for each time point t, $R(a,b,t)$ is the value of the relative risk that minimizes the number of observed cases that is needed to reject the null hypothesis of the classical SPRT with $\alpha = a$ and $\beta = b$. The test statistic is then defined as

$$A_t = \frac{P(c_t \mid HA : RR = R(a,b,t))}{P(c_t \mid H0)} = e^{(1-R(a,b,t))\mu_t} R(a,b,t)^{c_t}$$

The only difference with Wald's classical SPRT is that the pre-fixed value of r, which is constant over time t, is replaced by the function $R(a,b,t)$, which changes with t. In essence, in the beginning, the surveillance is conducted as if the alternative hypothesis has a high relative risk and, as the surveillance progresses, the relative risk in the alternative hypothesis decreases.

The upper and lower rejection limits are set to be $ln((1-b)/a)$ and $ln(b/(1-a)$ respectively, as with the classical SPRT. As Abt[16] points out, because of the minimization done when calculating $R(a,b,t)$, a and b no longer represent the approximate type 1 and 2 errors. Rather, for any pair (a,b), the true type 1 and 2 errors are calculated using computer simulations. For example, with $a = 0.07$ and $b = 0.08$, the type 1 error is $\alpha = 0.1$ and the type 2 error is $\beta = 0.05$.[16]

Maximized sequential probability ratio tests (MaxSPRT)

Developed in 2004 for the Vaccine Safety Datalink project, another approach to overcome the problem with Wald's classical SPRT is the maximized sequential probability ratio test (MaxSPRT),[17] where the alternative hypothesis is composite ($H_A : RR > 1$) rather than singular ($H_A : RR = r$).[18] In contrast to Abt's method, MaxSPRT defines the test statistic by maximizing the likelihood over different relative risk parameter values, which is the standard way to deal with composite alternative hypotheses.[19] There are different versions of the MaxSPRT depending on the nature of the data.

MaxSPRT: Poisson probability model

The Poisson-based MaxSPRT is used when observed counts are compared to expected counts. Its use is based on the likelihood ratio test statistic

$$LR_t = \max_{HA} \frac{P(c_t \mid HA)}{P(c_t \mid H0)} = \max_{HA:RR>1} \frac{e^{-RR\mu_t}(RR\mu_t)^{c_t}/c_t!}{e^{-\mu_t}(\mu_t)^{c_t}/c_t!}$$
$$= \max_{HA:RR>1} e^{(1-RR)\mu_t} RR^{c_t} = e^{(\mu_t - c_t)} \left(\frac{c_t}{\mu_t}\right)^{c_t}$$

The last equality holds since the maximum likelihood estimate of RR is c_t/μ_t. The equivalent but more computationally friendly log likelihood ratio test statistic is

$$LLR_t = \mu_t - c_t + c_t \ln(c_t/\mu_t)$$

Wald's classical SPRT rejects the null when the LLR reaches an upper bound, and accepts it when it reaches a lower bound, as determined by the mathematical formulas above. Upper and lower boundaries can also be used for the MaxSPRT, but there are also other options.

One possibility is to eliminate the lower boundary, and instead continue the surveillance until either the upper boundary is reached and the null hypothesis is rejected or until a predetermined maximum sample size is reached, defined in terms of the expected number of events accrued under the null hypothesis. For drug and vaccine safety surveillance this latter option is often more natural. Since such systems use observational surveillance

data that are collected for other purposes, there is no harm to patients if surveillance continues when the drug/vaccine is safe and the additional data analytic costs are minor.

A positive aspect of the Poisson-based MaxSPRT is that exact critical values can be calculated numerically. Users do not have to make the calculations themselves, as long as they use one of the maximum values on the sample size used in pre-calculated tables of critical values.[17] With a larger maximum sample size, the LLR-based critical values that define the upper boundary must be larger. This is natural since there is more multiple testing that needs to be adjusted for.

The standard Poisson-based MaxSPRT is based on the assumption that the expected number of cases is known under the null hypothesis. In practice, the expected counts are often estimated from a historical comparison group so that there is some uncertainty about them. If the sample size in the historical control group is at least five times larger than the predetermined maximum sample size for the surveillance, this is not a problem. If it is less, the uncertainty in the estimates will create a non-trivial bias, where the actual alpha level is greater than 0.06 for a nominal significance level of 0.05.[20] A conditional MaxSPRT should then be used instead.

Conditional MaxSPRT: Poisson probability model with a small control group

When there are no reliable estimates for the expected counts, one should use the conditional maximized sequential probability ratio test (CMaxSPRT), which is also based on the Poisson distribution.[20] With the CMaxSPRT, the randomness from both the historical data and the surveillance population is taken into account. To do so, the problem is envisioned from a different perspective. In both Wald's SPRT and the standard MaxSPRT, the number of adverse events is considered to be random while the cumulative number of people exposed to the vaccine is considered to be fixed. The CMaxSPRT does it the other way around, conditioning on the numbers of adverse events in both the control and surveillance populations, considering both of them to be fixed and non-random. The random variable is instead the number of vaccinated people that it takes to observe the observed number of adverse events. The intuitive idea is that if it takes fewer vaccinated individuals to observe the same number of adverse events in the surveillance population than in the historical data, then it is more likely that there exists some excess risk due to the vaccine.

Critical values, statistical power, and the expected time-until-a-signal have been calculated using computer simulations and can be obtain from existing tables.[20] Repeated confidence intervals[21] can also be calculated for the CMaxSPRT.[20] As with the standard Poisson-based MaxSPRT it is possible to adjust the CMaxSPRT for confounders such as age, gender, or study site.

MaxSPRT: binomial probability model

The binomial-based MaxSPRT is appropriate to use when comparing adverse event counts in some form of matched setting. In a matched concurrent control setting, the number of adverse events among users of a new drug or vaccine is compared to the number of adverse events among non-users that are matched by, for example, age, gender, time of vaccination, or chronic disease status. In a self-controlled setting, the number of an adverse event during an exposed time period is compared to the number of adverse events in an unexposed time period from the same set of individuals.

The binomial-based MaxSPRT uses the binomial probability model when calculating the likelihood ratio. Given that there is an adverse event in the data, it could either be for the exposed person or for an unexposed matched control. In essence, we have a number of coin tosses (adverse events) which may either turn up as a head or a tail (exposed or unexposed), and under the null hypothesis the probability of a head is known to be p ($p = 0.5$ for a 1:1 matching ratio, $p = 0.25$ for a 1:3 matching ratio, etc.). The upper limit on the length of surveillance is different for the binomial MaxSPRT, and should now be defined in terms of the observed number of adverse events. That is, the surveillance continues until there is a signal reject-

ing the null hypothesis or until the predetermined total number of adverse events is reached, in the exposed and unexposed time periods combined. Other than these differences, the principles behind the MaxSPRT are the same for Poisson and binomial type data, and the mathematical formulas for the binomial likelihood ratio test statistic are not given here.

Exact critical values for the binomial MaxSPRT can be calculated numerically, using an iterative Markov chain approach. A table with these critical values has been constructed for different fixed matching ratios, different length of surveillance, and different alpha levels.[17] Note that the matching ratio does not need to be an integer. For example, in a self-control setting, the exposed window could be 10 days long while the unexposed window is 25 days, leading to a 1 : 2.5 ratio.

The standard binomial MaxSPRT requires that the matching ratio is fixed over time. That is easily achieved for self-control designs. With concurrent matched controls, there may be a different number of suitable controls available for each exposed person, and there is then a loss of power if some suitable controls are not used. To solve this problem, Fireman *et al*. developed a modified and more flexible version of the binomial MaxSPRT, for which there can be a different number of matched controls for each exposed person, and where these matching ratios do not need to be known in advance (Fireman B, personal communication, 2010). Critical values can be calculated exactly, so again, there is no need for computer simulations.

Group sequential analysis methods

In contrast to continuous monitoring methods, where an analysis can be conducted as often as one wants, group sequential tests are performed at discrete time intervals after a certain amount of data has accumulated. In randomized clinical trials, group sequential designs are widely used to monitor the efficacy and safety of new medical products.[22–26] Sometimes they are used to reduce the expected sample size, shortening the expected length and cost of the study compared to a conventional fixed sample size design where one single analysis is performed at the end of the study. Most often though, they are used for ethical reasons so that a trial can be stopped if there are unsuspected and serious adverse events or if the efficacy of the treatment is so beneficial that it should not be withheld from the control group. Because of their use in clinical trials, group sequential methods have been studied much more than continuous sequential methods, and they offer considerable flexibility and variety in the types of hypotheses that can be addressed, the choice of test statistic used, the frequency at which the data are analyzed, and the shape of the signaling threshold over time.[25,26]

With some exceptions, group sequential methods have not yet been widely considered for safety surveillance that uses routinely collected observational data.[27–29] That is likely to change though (see Chapter 30). For logistical reasons, it is not always possible to obtain observational safety data on a near real-time basis, such as weekly or monthly. Instead, data may only arrive, for example, once a year. Continuous sequential methods would then adjust for too many analyses of the data, and it is then better to use a group sequential method.

Li[28] has proposed a Poisson-based conditional group sequential method to be used with a concurrent comparison group, such as a comparison drug in parallel usage. The method "imposes an unrestrictive semi-parametric Poisson regression model for the numbers of incident adverse events associated with" the two drugs, respectively, and "derives exact conditional inference for the parameter of interest based on the conditional distribution of our test statistics given the values of the sufficient statistics of the nuisance parameters."[28]

Hocine *et al*.[27] have proposed a group sequential case-series analysis for drug and vaccine safety surveillance, based on the self-controlled case series method. The analysis is based only on cases occurring within a certain age period, such as children age 12 to 24 months. This time period is then divided into exposed and unexposed intervals. The set-up is hence different from the self-control design described above. For example, adjustment for age is a key issue, and Hocine *et al*.[27] propose a few different options for such adjustments.

Selection of the comparison group for sequential analysis

As with any observational data, it is important to select an appropriate comparison group when doing prospective drug or vaccine safety surveillance, and compared with standard pharmacoepidemiologic studies, there are special considerations and restrictions due to the sequential nature of the analysis. For example, there are very little concurrent data at the start of surveillance.

The best choice of comparison group depends both on the medical product and on the potential adverse event under surveillance. The comparison group could be concurrent or historical; it could be users of a drug or vaccine that is being replaced, that is a competitor, which is of the same class or is used by approximately the same patient population. It also could be non-users defined through well-care visits or health plan membership. It is also possible to use self-controls, comparing exposed and unexposed time windows from the same individual.[11,30–34]

The comparison group should ideally be drawn from the same database, but could also be obtained from other data sources or the scientific literature.

Since different comparison groups are prone to different types of bias, it sometimes makes sense to run two or more different sequential analyses for the same drug–event pair, using different comparison groups.

We now describe some of the more common and important options for selecting the comparison group.

Population controls with confounder adjustment

Suppose we have a new drug or vaccine replacing an old one, such as the MMRV (measles, mumps, rubella, varicella) vaccine replacing the MMR (measles, mumps, rubella) vaccine, and that we want to test the null hypothesis that the risk of an adverse event after the new vaccine is the same as after the old one. Historical data can be used to calculate the number of adverse events after the old vaccine in various demographical population groups based on age, gender, study site, etc., which in turn is used to calculate the covariate-adjusted expected number of adverse events after the new vaccine. The covariate-adjusted expected counts can either be calculated using indirect standardization, if the covariates are categorical, or through a Poisson regression analysis. The Poisson-based MaxSPRT is an appropriate sequential method to use for this type of control group.

There is not always a specific drug/vaccine that is being replaced. One can then use another comparison drug/vaccine or well-care visits, instead. Another option is to use all unexposed days for all individuals in the data set, irrespectively of any other drug or vaccine utilization. For very rare diagnostic events, for which good estimates of the expected counts cannot be obtained from the same data base, it may be better to base the expected counts on incidence estimates obtained from the scientific literature.

The main advantage of a confounder-adjusted population-based comparison group is its typically large sample size. This is especially useful for rare adverse event outcomes.

No matter how the comparison group is selected, there is always a possibility of selection bias, due to some underlying differences in the disease risk with the exposed group. Another potential source of bias is seasonal and temporal trends, which may be due to temporal changes in disease incidence, the patient population, or medical coding practices.

One way to eliminate bias due to temporal trends is to use concurrent population-based controls, with a separate confounder adjustment performed for each time period. In order for the expected counts to be reliable, the size of the control population must be large so they have many more adverse events in each time period compared to the total number of adverse events in the exposed population during the whole length of the surveillance. This is easier to achieve for group sequential methods than for continuous sequential methods, since the concurrent control population will be larger when there is a longer time period between the repeated analyses performed.

Matched unexposed controls

In order to avoid bias due to secular or seasonal trends, one can also use a concurrent matched control design. For each exposed individual one or more unexposed persons are selected as controls, matched by age, gender, study site, geographical location, disease history, and time of exposure. The matching must occur in real time as the sequential analysis proceeds. The binomial MaxSPRT is a suitable method for this design, but it requires that the matching ratio is held constant over time. If there are a different number of matched controls per exposed individual, one should instead use the more flexible version proposed by Fireman *et al*.

The choice of population from which the matched unexposed controls are drawn is critical. The controls could, for example, be selected from children with a well-care visit during the same week as the exposed child was vaccinated; from people receiving another drug at the same time as the exposed person receives the drug under study; or they could be selected more generally from all health plan members irrespectively of their healthcare status. The best choice will depend on the drug/vaccine and the adverse event under surveillance. No matter how well the matching is done though, there is always a risk for selection bias.

It is also possible to use matched controls from a historical population. This could potentially decrease selection bias by making it easier to find suitable matched controls, but it could also increase selection bias depending on how the historical population is defined, at the same time as it increases the risk of bias due to seasonal or temporal trends.

One disadvantage of the matched control design, as well as the self-control design described below, is the inability to quickly generate a signal for very rare adverse events. For example, suppose that the background incidence rate of the adverse event is 1 per 1 000 000 annually, that exposure to a drug increases the risk 100-fold to 1 per 10 000, and that we are using a 1 : 1 matching ratio. After observing 40 000 exposed patient years, we would expect to see 4 adverse events in the exposed group and 0.04 events in the matched, unexposed group. In a Poisson-based analysis this is highly statistically significant, since the probability for this outcome is 0.0000000001. In a 1 : 1 matched control setting, the probability of observing 4 in the exposed and 0 in the unexposed group is (1/2)^4 = 0.0625, which is not statistically significant.[17] Hence, for rare adverse event outcomes one will have more power by using a population-based comparison group.

Self-controls

In a self-control design, the exposed time period right after the drug or vaccine is taken is compared to an unexposed time period from the same individual, either before the drug or vaccine is taken or long after the exposure is over. For example, the exposed time interval may be 1–14 days after vaccination while the unexposed time interval may be 15–28 days prior to the vaccination. For drugs, the time of vaccination is replaced by the day the drug was dispensed. The binomial MaxSPRT is suitable for this design, and can be used for any choice of time intervals.

There are both advantages and disadvantages with the self-control design. The main advantage is that there is no selection bias due to differences that do not change over time in the patient populations in the exposed and control groups, which is a concern with both the historical and concurrent matched control designs.

It does not adjust for time varying covariates though (i.e., factors that change in the pre- versus postcomparison period), so there is still the possibility of bias. For example, if the unexposed window is before the exposure, there may be confounding by indication or contraindication, if the adverse event makes a patient more or less likely to receive the vaccine or drug. For example, if a person that was just diagnosed with Guillain–Barré syndrome is less likely to receive the vaccine than the average person, that will artificially decrease the adverse event counts in the unexposed window before vaccination. The extent of this problem depends both on the drug/vaccine exposure and on the adverse event.

There could also be bias due to seasonal effects or temporal trends. For example, if we are interested in the safety of the influenza vaccine, the unexposed window is before the exposed window,

and there is a seasonal increase in the number of adverse event during the vaccination season, then there will be a bias towards signaling since there will be more adverse events in the later occurring window.

Age is another potential source of bias, since the patients are slightly older in the postvaccination exposure window than in the prevaccination unexposed window. For example, if the interest is in a vaccine for infants, a potential adverse event for which the natural incidence rate increases rapidly during the first few months of life, then such bias could generate a false signal. However, if the windows are short and close to each other, it should not be a significant source of bias except in rare situations.

Applications of sequential analysis in postmarketing safety surveillance

Vaccine Safety Datalink

Starting in 2004, the Centers for Disease Control and Prevention (CDC)-sponsored Vaccine Safety Datalink (VSD) project[2] began pioneering the use of sequential statistical methods for near real-time safety surveillance of new vaccines,[4] using weekly data from eight managed-care organizations. Using the MaxSPRT, seasonal influenza vaccine was monitored during the 2004/05 season. Subsequent vaccines for which surveillance has been performed in near real-time include a meningococcal conjugate vaccine for adolescents (MCV4; Menactra),[11] two tetanus–diphtheria–acellular pertussis vaccines for adolescents and adults respectively (Tdap; Adacel and Boostrix),[30] a combination measles–mumps–rubella–varicella vaccine for infants and children (MMRV; ProQuad),[33] a pentavalent bovine-derived rotavirus vaccine for infants (RotaTeq),[31] a human papillomavirus vaccine for young women (HPV; Gardasil), a combination vaccines for infants (DTaP-IPV-Hib; Pentacel), and seasonal and H1N1 influenza vaccines.[31,35,36]

For each vaccine, a different set of five to ten plausible adverse events are monitored, such as allergic reactions, Bell's palsy, encephalopathy/encephalitis, facial paralysis, fever, gastrointestinal bleeding, Guillain–Barré syndrome, intussusception, meningitis, seizures, and thrombocytopenia. The choice is based on biological plausibility, known safety problems with related vaccines, safety concerns from prelicensure clinical trials or other data sources, and a review of the existing literature.

In VSD, the most commonly used sequential study designs have been: (i) the Poisson-based MaxSPRT or CMaxSPRT, using historical VSD data to estimate expected counts from users of the old vaccine that the new vaccine is replacing, together with indirect standardization to adjust for age, gender, and health plan; (ii) the binomial-based MaxSPRT with a self-control design, comparing the number of adverse events during an exposure window immediately after the vaccination with an unexposed window either before or a long time after vaccination; and (iii) the binomial MaxSPRT or Fireman *et al.*'s more flexible version, using concurrent controls who came in for a well-care visit, matched by age, gender, health plan, and the approximate week of vaccination.

Yih *et al.*[37] have reviewed the first 4 years of sequential analyses in the VSD. Some statistical signals were explained by data errors, confounding, or chance, and with one exception (discussed below), no actual safety problems were detected. It is still very important to conduct such surveillance, as it provides added reassurance about vaccine safety to patients, parents, and physicians.

For the MMRV vaccine, weekly sequential analyses detected an increased risk of seizures among infants compared to users of the MMR vaccine.[33] This statistical signal was investigated using temporal scan statistics and logistic regression and confirmed to be real. As a result, the Advisory Committee on Immunization Practices (ACIP) modified their vaccine recommendations in February, 2008, no longer recommending MMRV over separate vaccinations of MMR and varicella, and the Food and Drug Administration and the manufacturer revised the product label.[33,38] In subsequent pharmacoepidemiologic studies, this increased risk was confirmed when compared to separate MMR and varicella vaccinations, and estimated to be about one additional seizure for every 2500 vaccine doses, occurring 7–10 day after vaccination.[33,39]

Other vaccine safety surveillance

Following the lead of the Vaccine Safety Datalink, there are now several other systems that use sequential analysis for near real-time postmarketing vaccine safety surveillance. During the 2009/10 influenza season, weekly sequential analyses were performed for seasonal and H1N1 influenza vaccines in the FDA-funded and CDC-sponsored Post-Licensure Rapid Immunization Safety Monitoring (PRISM) surveillance system, using data from several health insurance plans and state immunization registries, covering more than 10% of the United States population. This system used the Poisson- and binomial-based MaxSPRT, as well as the CMaxSPRT. Other near real-time H1N1 safety surveillance systems have used data from Medicare,[40] the Indian Health Service,[41] and a private health insurance plan,[42] using either continuous[41,42] or group[40] sequential methods.

Sequential methods can also be used for data from spontaneous reporting systems. For such data, the Medicines and Healthcare products Regulatory Agency in the United Kingdom have used the Poisson-based MaxSPRT for near real-time safety surveillance for adjuvant H1N1[43] and HPV vaccines.[44]

Drug safety surveillance

Sequential analysis has also be considered for near real-time postmarketing surveillance to evaluate the safety of pharmaceutical drugs. Drug safety surveillance is more complex than vaccine safety surveillance for three main reasons: (i) people are exposed to a drug over a longer time period, as opposed to a vaccination that occurs on a known single day; (ii) we know almost surely that a person actually got the vaccine as it is administered by a health-care worker, while drug exposure is usually estimated from dispensing or prescribing data without certain knowledge about if and how often the drug was actually taken; and (iii) most vaccines are given to a high proportion of a mostly healthy population, such as infants, while a drug is given to a selected patient populations, such as diabetics, which may have pre-existing health conditions that could be related to the adverse event under surveillance.

Brown *et al*.[45,46] demonstrated the promise of using sequential methods with health insurance claims data to construct a system for the early identification of adverse drug events (ADEs). They assessed the ability to detect ADEs using historical data from nine health plans involved in the HMO Research Network's Center for Education and Research on Therapeutics (CERT). Analyses were performed using the Poisson-based MaxSPRT. Five drug–event pairs representing known associations with an ADE and two pairs representing 'negative controls' were analyzed. Statistically significant ($p < 0.05$) signals of excess risk were found in four of the five drug–event pairs representing known associations; no signals were found for the two negative controls. Signals were detected between 13 and 39 months after the start of surveillance. More recently, Wahl *et al*.[47] have evaluated the use of sequential methods for drug safety surveillance as one part of a comprehensive claims-based Healthcare Safety Surveillance System (HSSS).

In the first efforts to sequentially analyze data obtained in near real-time, Avery *et al*.[48] used the MaxSPRT to compare the safety of a generic anticonvulsant (divalproex sodium) compared with the branded product. Chen *et al*.[49] used the MaxSPRT to monitor the risk of infections among patients treated with tumor necrosis factor (TNF) inhibitors.

Statistical design considerations

Whether one uses continuous or group sequential methods, whether one uses a Poisson or binomial probability model, and independently of the comparison group used, there are several important design consideration that are specific to sequential statistical analyses. These determine the probability to reject the null when it is true, the statistical power to detect a true signal, the time until a signal is detected, and the length of the surveillance. Here are the three main statistical design issues, and a discussion on how they differ between postmarketing safety studies and clinical trials, where sequential statistical methods have been most commonly used to date.

Sample size

In a standard non-sequential analysis, sample size and statistical power go hand in hand, with the latter increasing as the former increases. For sequential analyses, the same general principle obviously holds, but it is not as straightforward. First of all, there is the (i) rate at which the sample size accrues over time, such as 1000 new vaccinees per week or 100 000 new drug exposed person-days per month, (ii) the final accrued sample size when the surveillance ends, and (iii) the maximum sample size at which the surveillance will end, no matter what. All of these are important concepts, since we are just as interested in the time until a signal occurs as in the overall statistical power. For example, by doubling the size of the population under surveillance, we double the rate at which the sample size accrues, cutting the time until a signal occurs in half, even if the maximum sample size is unchanged.

In a clinical trial, each observation is costly, and there is often a budgetary restriction on the maximum sample size at the end of the study, but the rate of sample size accrual is often determined more by logistical than financial consideration. For an observational safety study, it is often costly to increase the rate at which the sample size accrues, as one would have to add additional data sets such as another health plan, but the cost to continue surveillance for a little longer is typically very small in a surveillance system that is already in place. Moreover, if there is a slight delay in detecting a safety problem in a Phase III clinical trial, it only affects the few individuals in the study. In postmarketing safety surveillance, there are many people exposed to the drug or vaccine that are not part of the surveillance system, and a slight delay in detecting a safety signal can affect many people.

Hence, in postmarketing safety surveillance, it usually is relatively easy and cheap to increase the overall statistical power of the surveillance system, but it is costly to reduce the time until a signal occurs, which is of great importance. What this means is the following. Once the population under surveillance has been chosen and the desired overall statistical power to detect a clinically important excess risk has been defined, one can almost always select the required maximum sample size to achieve that power. On the other hand, if the expected time until a clinically important signal occurs is too long, one must increase the rate at which the sample size accrues by adding additional datasets to increase the size of the population that is under surveillance. Hence, with respect to sample size, the most important trade-off is between the time to a signal and the sample size accrual rate.

Stopping boundary shapes

Another sequential design feature that influences overall statistical power and signal detection timeliness is the choice of the stopping boundary shape. Wald's classical SPRT and the maximized SPRT both use a flat upper boundary with respect to the log likelihood ratio (LLR), such that the null hypothesis is rejected when the LLR exceeds a certain fixed critical value. Wald's SPRT surveillance ends without rejecting the null when the LLR reaches a lower flat boundary, while the MaxSPRT ends surveillance without rejecting the null when the maximum sample size or the maximum number of adverse events is reached. This is a natural choice for drug and vaccine safety applications but not the only option. The MaxSPRT can be used with any other type of rejection bounds as well, including the traditional upper and lower bounds used by the classical SPRT as well as the boundaries proposed by Pocock,[50] O'Brien and Fleming,[51] and others that are commonly used for group sequential methods in clinical trials.

If the maximum sample size is fixed, overall statistical power is maximized by doing one single non-sequential analysis at the end of the study. The use of a sequential method will then always reduce the overall power, but the amount of reduction depends on the boundary used. Boundaries like the O'Brien-Fleming boundary that are high in the beginning, making it more difficult to reject the null hypothesis early on during the surveillance, will by default have higher overall statistical power, although they will often take longer to generate a signal. Hence, when the maximum sample size is fixed, there is a trade-off between overall power and timeliness when selecting the stopping boundary.

In a clinical trial, overall power is of primary importance. Within a restraint of the maximum sample size, the goal is often to lose as little overall power as possible while still having some ability to stop the trial early if a serious safety problem arises. In clinical trials where a new drug or procedure is tested, a slight delay in the time-to-signal will only affect a very small number of people, since it is not yet in general circulation. Timeliness to signal is often, therefore, of secondary importance.

In contrast, in postmarket safety surveillance it is usually timeliness that is of primary importance. Since the drug or procedure is already in general use, many adverse events can potentially be prevented in the general population if a safety signal is generated a few weeks or months earlier. Overall power is of course also important, but it is often relatively easy to achieve at very limited cost by simply increasing the maximum sample size. Rather than the trade-off between overall power and timeliness-to-signals that exist in the clinical trial setting, for postmarketing safety surveillance, the choice of stopping boundary is primarily a trade-off between timeliness-to-signal for modest versus high excess risks. How different choices of the stopping boundary affect this trade-off has not yet been fully investigated.

Frequency of the analyses

The frequency of analyses in a sequential setting is primarily determined by how frequently data can be obtained from the health plans or other data providers. It is important to note though that different data providers do not need to be synchronized. One health plan may provide data every Monday morning for the events that happened during the prior week while another health plan may only be able to provide data once a month with a 10-week delay. With a sequential statistical design, the data can be analyzed as it arrives to the data coordinating center independently of when the events happened and from where it arrived.

With a continuous sequential method, the data can be analyzed as often as one likes, without having to worry about multiple testing. If the data are analyzed at discrete time points but still frequently, such as once a week, a continuous sequential method will be only slightly conservative so that the true alpha level is slightly higher than the nominal one. If analyses are done less frequently, such as once or twice a year, the true and nominal alpha levels may differ greatly, and it is then better to use a group sequential method. Research has not yet determined where the cut-off is. First of all, it depends on the amount of new data received between each data arrival rather than on the number of days, weeks, or months between such arrivals. Second, many group sequential methods are approximate in nature, based on either asymptotic theory or computer simulations. Since the critical values of both the Poisson and binomial MaxSPRT are based on exact calculations, they can be viewed as one way to approximate the critical values when the data arrive in frequent intervals, with the advantage that the bias always goes in the conservative direction.

The frequency at which new data arrive is often a design consideration that can be controlled by the investigators. Sooner and more frequent is always better than later and less frequent from an analytical and performance perspective, but the former may be more costly in terms of both money and effort. It is then a trade-off between cost of setting up the surveillance system and its timeliness to detect a safety problem.

Signal investigation

The purpose of near real-time sequential analysis is to generate alerts about potential vaccine or drug safety problems. However, such surveillance is only the first step, and when there is a statistical signal, it is important to conduct further investigations and analyses before making definite conclusions or reporting results. There are several reasons for such caution:

- Electronic health data are collected for clinical and administrative purposes and they occasionally contain inaccuracies. When data arrive and are analyzed in near real time, it is not possible to thoroughly check all the data before the analyses are run.
- Prospective surveillance systems are usually conducted simultaneously for many drug–event pairs. It may then be more convenient and efficient to

only adjust for the common and most important covariates as part of the sequential analyses, such as age, gender, seasonality, and study site, while the remaining drug–event-specific confounders are adjusted for in subsequent signal investigation studies, as needed.

• In observational studies, there are many potential sources of bias, and all of them cannot be accounted for in a single sequential analysis. While multiple sequential analyses are sometimes run in parallel for the same drug–event pair, with different comparison groups, many detailed adjustments and sensitivity analyses are more conveniently left for the signal investigation phase.

There are several approaches and methods available for investigating a statistical signal. Although a full description is beyond the scope of this chapter and further details are provided throughout this book, here we simply give a brief, general description of some of the most important and commonly used steps,[11,30,31,33,37] in approximate but not strict chronological order:

1 Check the basic data, including observed and expected counts. Compare with incidence and prevalence estimates from the literature.

2 Tabulate descriptive statistics by age, gender, and study site. Compare drug utilization and event rates at the different study sites. Look at secular and seasonal trends.

3 Check the computer code, including both the analysis code and the data generation code at each study site. Use the same data to do an equivalent non-sequential analysis, which should give similar risk estimates.

4 Look at the time from initial drug exposure to the adverse event, using descriptive histograms. If there is no relationship between the drug and diagnostic event, these should be roughly uniformly distributed. Consider different risk windows. Formal statistical inference can be done using the temporal scan statistic,[52,53] which adjusts for the multiple testing inherent in the many different potential risk windows evaluated.

5 Adjust for a different and larger set of potential confounders using standard non-sequential pharmacoepidemiologic methods on the same data set. This may include more detailed age adjustments, adjustments for seasonal trends using months or sinusoidal curves, adjustments for secular trends, adjustments of day-of-the-week effects, adjustments for chronic disease conditions, adjustments for concomitant vaccines or medications, etc.

6 Use other, different comparison groups than the one used in the sequential analysis. This may include historical comparison groups from different time periods, different definitions of matched controls, and the use of different time periods in self-control designs.

7 Conduct a chart review of adverse events to exclude erroneously diagnosed cases from automated electronic health records. This could be a complete review, or a review of only a random subsample of the exposed and/or unexposed individuals. Redo the non-sequential analyses with only the chart-confirmed individuals.

8 Compare the signal generated by one drug–event pair with results for subdiagnostic groups and with results from similar drugs and diagnostic events. For example, if there is a signal indicating an increased risk of febrile seizures, check if there is also an increased risk of fever, even if that by itself would not be of interest.

9 Compare the results with other existing data sets, such as: Phase III clinical trials; Phase IV postmarketing trials; spontaneous adverse event reporting systems such as AERS and VAERS; and other observational data sets such as electronic health records from a different health plan.

10 Collect more data. Continue the prospective monitoring of a drug–event pair even after a statistical signal has been generated, to see if the effect size increases or decreases over time. Conduct a completely new study designed from scratch, such as a case–control study or a postmarketing randomized trial.

The future

Sequential statistical methods have proved successful for prospective near real-time postmarketing

vaccine safety surveillance, and newly approved vaccines are now routinely monitored for adverse events using weekly data.[37] Efforts are underway to establish similar systems for postmarketing safety surveillance for pharmaceutical drugs (see Part III and Chapter 30).[45–49] While it is too early to tell how successful drug safety surveillance systems will be, there is some reason for optimism despite the more complex nature of the data.

Future research needs

The use of sequential statistical methods for postmarketing safety surveillance is still in its infancy, and there are many statistical, epidemiologic, and logistical questions to be answered before we will know how to best design near real-time safety surveillance systems. Sequential methods have primarily been used in randomized clinical trials where the data and the objectives are very different, so conclusions from that literature do not necessarily carry over to observational safety surveillance. More research is needed on a wide range of sequential design options such as the sample size, boundary shape, and testing frequency so that trade-offs between overall statistical power, time-to-signal, and the duration of surveillance are better understood. Another issue that needs further evaluation is the trade-off between timeliness to signals and the logistical and financial implications from different testing frequencies.

There are also more epidemiologic-oriented research needs for near real-time surveillance.[34] What is the best choice of comparison group in different settings? How can we better adjust for confounding? What is the best way to deal with missing data? How do we best analyze safety problems (e.g., thrombocytopenia) that are most naturally expressed using a continuous variable? What are the best ways to define drug exposure windows and diagnostic events of interest? Much of pharmacoepidemiologic methods research is focused on identifying better approaches to these problems, and sequential statistical analyses introduces unique constraints that require novel solutions to some of these issues.

Acknowledgements

For many stimulating methodologic discussions on sequential statistical methods, I wish to acknowledge my fellow VSD biostatisticians, Bruce Fireman, Margarette Kolczak, Ned Lewis, Lingling Li, David McClure, and Jennifer Nelson, who have all played an important role in the development and implementation of sequential analysis for postmarketing safety surveillance. The writing of this chapter was supported by funding from the Centers for Disease Control and Prevention, via America's Health Insurance Plans, for the Vaccine Safety Datalink project, contract number 200-2002-00732. The opinions and conclusions are the author's and do not necessarily represent the views or policies of the United States Department of Health and Human Services or America's Health Insurance Plans.

References

1. Varricchio F, Iskander J, Destefano F, Ball R, Pless R, Braun MM, *et al*. Understanding vaccine safety information from the Vaccine Adverse Event Reporting System. *Pediatr Infect Dis J* 2004; **23**: 287–94.
2. Chen RT, DeStefano F, Davis RL, Jackson LA, Thompson RS, Mullooly JP, *et al*. The Vaccine Safety Datalink: Immunization research in health maintenance organizations in the USA. *Bull World Health Organ* 2000; **78**: 186–94.
3. 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.
4. Food and Drug Administration. *The Future of Drug Safety—Promoting and Protecting the Health of the Public. FDA's Response to the Institute of Medicine's 2006 Report*, Jan 2007. http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM171627.pdf (accessed July 2011).
5. 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**: 483–6.
6. Moore N, Kreft-Jais C, Haramburu F, Noblet C, Andrejak M, Ollagnier M, *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**: 513–18.

7. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, *et al*. A Bayesian neural network method for adverse drug reaction signal generation. *Am J Clin Pharmacol* 1998; **54**: 315–21.
8. DuMouchel W. Bayesian data mining in large frequency tables with an application to the FDA spontaneous reporting system. *Am Stat* 1993; **53**: 177–90.
9. 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**: 381–92.
10. Davis RL, Kolczak M, Lewis E, Nordin J, Goodman M, Shay DK, *et al*. Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology* 2005; **16**: 336–41.
11. Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K, *et al*. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 2007; **45**: 89–95.
12. Wald A. Sequential tests of statistical hypotheses. *Ann Math Stat* 1945; **16**: 117–86.
13. Wald A. *Sequential Analysis*. New York: Wiley, 1947.
14. Page ES. Continuous Inspection Schemes. *Biometrika* 1954; **41**: 100–15.
15. Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Stat Methods Med Res* 2003; **12**: 147–70.
16. Abt K. Poisson sequential sampling modified towards maximal safety in adverse event monitoring. *Biometrical J* 1998; **40**: 21–41.
17. Kulldorff M, Davis R, Kolczak M, Lewis E, Lieu T, Platt R. A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Seq Anal* 2011; **30**: 58–78.
18. Lorden G. Open-ended tests for Koopman-Darmois Families. *Ann Statistics* 1973; **1**: 633–43.
19. Lehmann EL. *Testing Statistical Hypotheses*, 2nd edn. New York: Springer-Verlag, 1986.
20. Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. *Stat Med* 2010; **29**: 284–95.
21. Jennison C, Turnbull BW. Repeated confidence intervals for group sequential clinical trials. *Control Clin Trials* 1984; **5**: 33–45.
22. Jennison C, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, FL: CRC Press, 2000.
23. Whitehead J. *The Design and Analysis of Sequential Clinical Trials*. New York: Wiley, 1997.
24. Armitage P. Sequential methods in clinical trials. *Am J Pub Health* 1958; **48**: 1395–402.
25. Emerson S, Kittelson J, Gillen D. Tutorial in biostatistics: Frequentist evaluation of group sequential clinical trial designs. *Stat Med* 2007; **26**: 5047–80.
26. Emerson S, Kittelson J, Gillen D. Bayesian evaluation of group sequential clinical trial designs. *Stat Med* 2007; **26**: 1431–49.
27. Hocine MN, Musonda P, Andrews NJ, Farrington CP. Sequential case series analysis for pharmacovigilance. *J Royal Statistical Society A*, 2009; **172**: 213–36.
28. Li L. A conditional sequential sampling procedure for drug safety surveillance. *Stat Med* 2009; **28**: 3124–38.
29. Shih MC, Lai TL, Heyse JF, Chen J. Sequential generalized likelihood ratio tests for vaccine safety evaluation. *Stat Med* 2010; **29**: 2698–708.
30. Yih WK, Nordin JD, Kulldorff M, Lewis E, Lieu TA, Shi P, *et al*. An assessment of the safety of adolescent and adult tetanus-diphtheria-acellular pertussis (Tdap) vaccine, using active surveillance for adverse events in the Vaccine Safety Datalink. *Vaccine* 2009; **27**: 4257–62.
31. Belongia EA, Irving SA, Shui IM, Kulldorff M, Lewis E, Yin R, *et al*. Real-time surveillance to assess risk of intussusception and other adverse events after pentavalent, bovine derived rotavirus vaccine. *Pediatr Infect Dis J* 2010; **29**: 1–5.
32. Greene SK, Kulldorff M, Lewis EM, Li R, Yin R, Weintraub ES, *et al*. Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project. *Am J Epidemiol* 2010; **171**: 177–88.
33. Klein N, Fireman B, Yih WK, Lewis E, Kulldorff M, Ray P, *et al*. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. *Pediatrics* 2010; **126**: e1–8.
34. 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.
35. Green SK, Kulldorff M, Yin R, Yih WK, Lieu TA, Weintraub ES, *et al*. Near real-time vaccine safety surveillance with partially accrued data. *Pharmacoepidemiol Drug Saf* 2011; **20**: 583–90.

36. Lee GM, Greene SK, Weintraub ES, Baggs J, Kulldorff M, Fireman BH, *et al*. on behalf of the Vaccine Safety Datalink Project. H1N1 and seasonal influenza vaccine in the vaccine safety datalink project. *Am J Prevent Med* 2011; **41**: 121–8.
37. Yih WK, Kulldorff M, Fireman BH, Shui IM, Lewis EM, Klein NP, *et al*. Active Surveillance for adverse events: the experience of the Vaccine Safety Datalink Project. *Pediatrics* 2011; **127**: S54–64.
38. Centers for Disease Control and Prevention (CDC), Advisory Committee on Immunization Practices (ACIP). Update: recommendations from the Advisory Committee on Immunization Practices (ACIP) regarding administration of combination MMRV vaccine. *MMWR Morb Mortal Wkly Rep* 2008; **57**: 258–60.
39. Jacobsen SJ, Ackerson BK, Sy LS, Tran TN, Jones TL, Yao JF, *et al*. Observational safety study of febrile convulsion following first dose MMRV vaccination in a managed care setting. *Vaccine* 2009; **27**: 4656–61.
40. Burwen D, Sandhu S, MaCurdy T, Kelman J, Gibbs J, Garcia B, *et al*. Surveillance for Guillain–Barre syndrome after influenza vaccination among the Medicare population, 2009–2010. *Pharmacoepidemiol Drug Saf* 2010; **19**: S15.
41. Qiang Y, Groom A, Redd J, Haberling D, Holman R, Jim C, *et al*. Safety of the influenza A (H1N1) 2009 monovalent vaccine among the Indian health service user population. *Pharmacoepidemiol Drug Saf* 2010; **19**: S329.
42. Enger C, Turnbull B, Gately R, Loughlin J, Chan KA. H1N1 influenza vaccine surveillance project. *Pharmacoepidemiol Drug Saf* 2010; **19**: S14–15.
43. Seabroke S, Bryan P, Wise L. The use of observed vs expected analyses in real-time vaccine safety surveillance in the UK during the H1N1 pandemic. *Pharmacoepidemiol Drug Saf* 2010; **19**: S188.
44. Seabroke S. Observed vs expected methodologies in vaccine safety surveillance: practical experience of using observed vs expected analyses for vaccine pharmacovigilance at the UK MHRA. *Presentation at the 26th International Conference on Pharmacoepidemiology and Therapeutic Risk Management*, Brighton, UK, 2010.
45. Brown JS, Kulldorff M, Chan KA, Davis RL, Graham D, Pettus PT, *et al*. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1275–84.
46. Brown JS, Kulldorff M, Petronis KR, Reynolds R, Chan KA, Davis RL, *et al*. Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations. *Pharmacoepidemiol Drug Saf* 2009; **18**: 226–34.
47. Wahl PM, Schneeweiss S, Wasser TE, Eisenberg DF, Rodgers K, Gagne JJ, *et al*. Development of a claims-based healthcare safety sentinel system featuring signal strengthening during sequential monitoring. *Pharmacoepidemiol Drug Saf* 2010; **19**: S80.
48. Avery TR, Vilk Y, Kulldorff M, Li L, Cheetham TC, Dublin S, *et al*. Prospective, active surveillance for medication safety in population-based health networks: a pilot study. *Pharmacoepidemiol Drug Saf* 2010; **19**: S304.
49. Chen Y, Strombom I, Marmaduke D, Benichou O, Lipkovich I. The risk of infections among patients treated with TNFi: population-based continuous ADRs monitoring using the maximized sequential probability ratio testing. *Pharmacoepidemiol Drug Saf* 2010; **19**: S5.
50. Pocock S. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; **64**: 191–9.
51. O'Brien P, Fleming T. A multiple testing procedure for clinical trials. *Biometrics* 1979, **35**: 549–56.
52. Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods* 1997; **26**: 1481–96.
53. Kulldorff M and Information Management Systems Inc. *SaTScan: Software for the Spatial and Space-time Scan Statistics*, v9.0: 2010. http://www.satscan.org (accessed July 2011).

CHAPTER 47

Advanced Approaches to Controlling Confounding in Pharmacoepidemiologic Studies

Sebastian Schneeweiss[1] *and Samy Suissa*[2]

[1] Division of Pharmacoepidemiology, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA

[2] Centre for Clinical Epidemiology, Lady Davis Research Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada

Introduction

The past two decades have witnessed an explosion of methodologic advances in the design and analysis of epidemiologic studies. Several 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 certainly take a 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.

We start out by describing a robust study design that will address 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 applications and others that only recently have made inroads into pharmacoepidemiology.

Clinical problems to be addressed by pharmacoepidemiologic research

Pharmacoepidemiologic analyses are in principle not different from analyses in any other subject area within epidemiology. They are concerned with valid estimation of associations between an exposure and outcome, struggling with 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 health-care databases from insurance health plans, electronic medical records systems, or registries (see Chapters 11–18 and 21). 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 37).[1] 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.

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

Methodologic problems to be addressed by pharmacoepidemiologic research

The ready and relatively cheap availability of large, longitudinal, patient-level health-care databases make the new-user cohort design a natural design choice as a starting point that mimics the classical parallel group controlled trial, except of course for the randomized treatment assignment (Figure 47.1).[2] Efficient sampling designs within such cohorts, including case–control, case–cohort, and two-stage sampling designs, are important extensions to assess additional covariate information in a subset of patients. Such sampling usually provides no advantage if secondary data are the only source for exposure and covariate assessment because there is no additional cost or time to analyze the entire dataset rather than a subsample.[3]

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 47.2). 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 below). Another option is random allocation of treatments between different patients. For most pharmacoepidemiologic studies, we will utilize variation in exposure among 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]

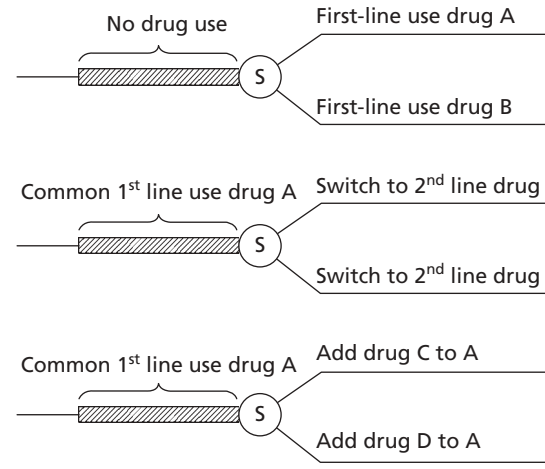


Figure 47.1 Principle of the new user design and its variations when studying second-line therapies. Reproduced from Schneeweiss[3] with permission from Wiley-Blackwell.

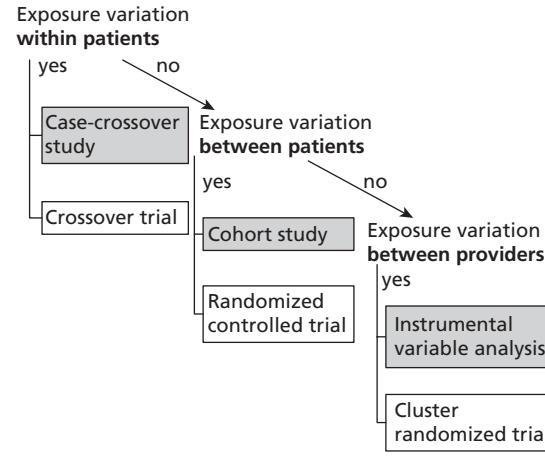


Figure 47.2 Study design choice by source of exposure variation. Reproduced from Schneeweiss[3] with permission from Wiley-Blackwell.

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. The clear temporal sequence of confounder adjustment 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] Identifying two active treatment groups further reduces the chances of immortal time bias, a mistake that most frequently emerges if future records are used to define earlier exposure status in health-care databases, particularly when defining a "non-user" comparison group.[6] A common example of immortal time bias is to define non-users as patients who have not used the study medication during the first 6 months of follow-up. By definition these non-user patients cannot die during the first 6 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 a longer time 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 (Figure 47.1).

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: (i) 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 (ii) 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 infeasible, particularly if the cohort size and number of outcome events are large. 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 multitime case–control, and case–cohort designs.

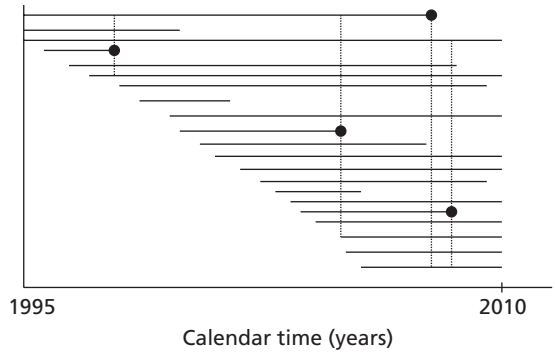


Figure 47.3 Illustration of a calendar-time cohort of 21 subjects followed from 1995 to 2010 with four cases (●) occurring and related risk-sets (—).

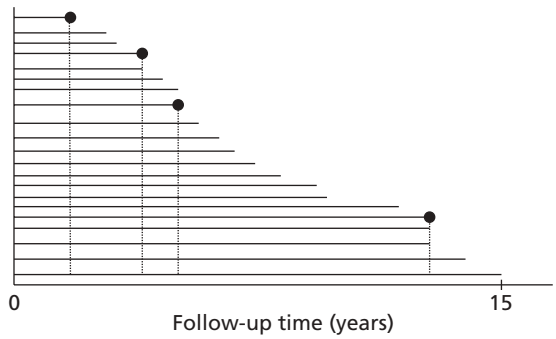


Figure 47.4 Illustration of follow-up-time cohort representation after rearranging the cohort in Figure 47.1 with the new risk-sets (—) for the four cases.

Structures of cohorts

Figure 47.3 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 exposure to a drug), which may be more relevant to the drug effect under study. In this instance, the illustration given in Figure 47.4 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 47.3 and 47.4 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 47.3, 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 47.4, 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.

The nested case–control design

The notion of a nested case–control design within a cohort was first introduced by Mantel,[10] who proposed an unmatched selection of controls and called it a synthetic retrospective study. It was developed further and formalized by Liddell *et al.*[11] 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 47.5 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 is introduced in the estimation of the relative risk if controls are forced to be selected only from the non-cases 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.[12] 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.

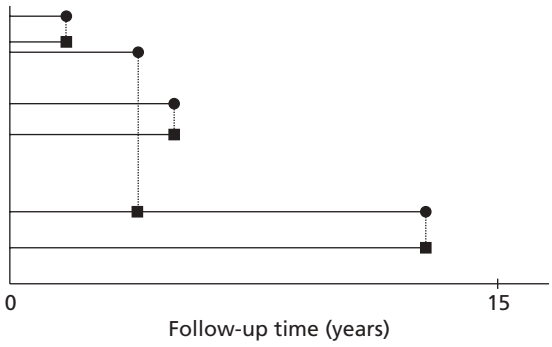


Figure 47.5 Nested case–control sample of one control (■) per case (●) from the cohort in Figure 47.4.

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.[13] A cohort of 12 301 asthmatics spanning the period 1978–1987 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 2-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 2-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 con-

tained 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 over 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.[14] Simple formulae exist to estimate the relative risk for 1:1 matching.[15] When more than one control is matched to each case, however, statistical packages such as SAS, STATA, or R are required to fit the necessary conditional logistic regression model.

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,[16] 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).[17,18]

Like the cohort, the nested case–control design is used primarily to conduct internal comparisons (within the cohort) between exposures to different drugs. At times, however, it is of interest to contrast exposure to drugs versus no exposure. This is not possible using methods for internal comparisons when all subjects in the cohort are exposed to the drugs under study. Instead, external comparisons are performed, comparing the rate of outcome in the cohort to that of an external population, with appropriate adjustment for only a few available ey factors, such as age, sex, and calendar time. The resulting measure is usually called the standardized mortality rate (SMR) or the standardized incidence rate (SIR). Techniques to estimate these measures using the full cohort are described in most textbooks of epidemiology.[15] Such techniques are not appropriate, however, in a nested case–control design, because subjects are not a simple random sample from the cohort. Indeed, it is evident from Figure 47.3 that cohort members with the longest follow-up have a greater chance of being selected in the nested case–control sample, since they belong to all the risk-sets. The appropriate method to perform external comparisons using data from a nested case–control design has been described.[19] 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.

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.[20] While the study collected data on exposure for 4 weeks prior to the index date, only 1-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 4 weeks (i.e., four control person-moments) rather than only 1-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.[21] 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 1 and 10 controls per case were identified.[21] The rate ratio of AMI associated with use of antibiotics in the month prior to the index date was 2.00 (95% 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 1 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,[22] in performing external comparisons of leukemia rates in patients treated by radiation for cervical cancer. It was ultimately developed and formalized by Prentice,[23] who coined the name "case–cohort." Although recent, this design has already been used effectively in some drug risk studies.[24–27] 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 47.6 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

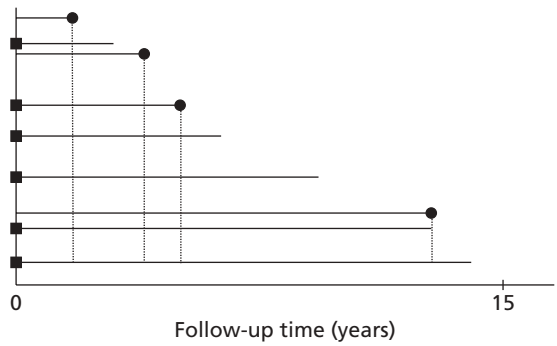


Figure 47.6 Case–cohort sample with six controls (■) from the cohort in Figure 47.4.

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 method of analysis for a case–cohort sample is complex, as it must take into account the overlap of cohort members between successive risk-sets induced by this sampling strategy. This handicap has severely limited the use of the case–cohort design. However, a statistical software package called EPICURE[28] has been released, which includes a module for the analysis of case–cohort data. Simplifications to the analysis have been developed along with SAS programs.[29] These advances will facilitate and undoubtedly encourage the future use of the case–cohort design, which offers some interesting advantages over the nested case–control design.[28]

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. In contrast, the nested case–control design requires different control groups for each type of event because the selection 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.[13] 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.[30]

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 larger sample sizes and 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.[31]

Within-subject designs

When dealing with the study of transient drug effects on the risk of acute adverse events, Maclure[32] 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. 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.

The 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 non-exposure 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 4 hours in our example. This occurrence is then classified as having arisen either under drug exposure or non-exposure 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 4-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.[32]

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 4 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 4 hours and on average over the past year. The data are displayed in Table 47.1. 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 4-hour periods being 2190 in 1 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 4-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 2 hours, then the data of Table 47.1 would be affected in two ways: some cases may not be considered exposed anymore, and the exposure probabilities will change. By considering as unexposed cases 2 and 4, for instance, who may have been exposed 3 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 6 hours long, then the data of Table 47.1 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 5 hours before ventricular tachycardia, and recalculating the appropriate exposure

Table 47.1 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 4-hour period after beta-agonist exposure

Case number	beta-agonist use in last 4 hours* (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

*Inhalations of 200 μg: 1 = yes, 0 = no.

Note: rate ratio estimator is $(\sum E_i N_{0i}) / (\sum (1 - E_i) N_{1i})$

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 eliminates 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 can be severe. However, 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.[33] There were 5579 cases of MVC identified from the Province of

Table 47.2 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

	Cases		Controls		OR	95% CI
	high	low	high	low		
Current beta-agonist use (case–control)	93	36	241	414	3.1*	1.8–5.4
Discordant† use (case-crossover)	29	9			3.2	1.5–6.8
Discordant† use (control-crossover)			65	25	2.6	1.6–4.1
Case-time-control	29	9	65	25	1.2	0.5–3.0

*Adjusted estimate from case–control analysis.
†Discordant from exposure level during reference time period.

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.[34–38] A cohort version has also been adapted for application to the risk assessment of vaccines (see Chapter 46).[39]

The 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.[40] 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,[13] 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 47.2 displays the data comparing low (12 or less 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 1-year current period and the 1-year reference period prior to the current period. First, a 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 excludes the effect of unmeasured confounding by disease severity, indicates a minimal 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.[41,42] This approach has been used successfully in several studies.[43–48]

Analytic approaches for improved confounding control

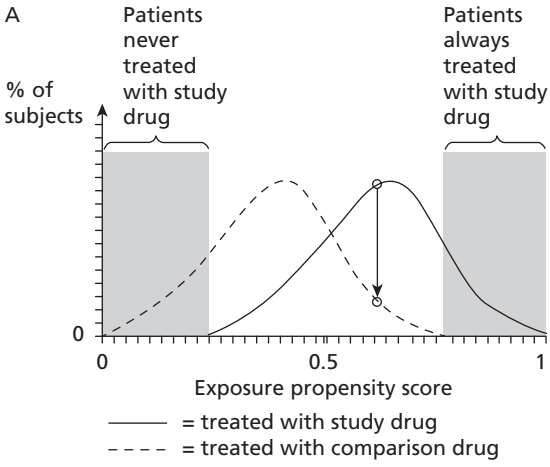
Balancing patient characteristics

Confounding caused by imbalance of patient risk factors between treatment groups is a known threat to validity in non-randomized studies of treatment effects. A litany of options for reducing confounding is available to epidemiologists.[49,50] Several approaches fit key characteristics of longitudinal health-care 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 pre-existing 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.[51] Estimating the propensity score using logistic regression is uncomplicated. Strategies for variable selection are well described.[52] 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.[53] Selecting variables based on *p*-values is not helpful as this strategy dependents 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 covariates there are several options to utilize it in a second step to adjust confounding. Typical strategies include adjustment from quintiles or deciles of the score with or without trimming, regression modeling of the PS, or matching on propensity scores.[54] Matching illustrates the working of propensity scores well.

Fixed ratio matching on propensity scores like 1 : 1 or 1 : 4 matching has several advantages that may outweigh its drawback of not utilizing the full dataset in situations where not all eligible patients match. 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 47.7). These tails of the PS distribution often harbor extreme patient scenarios that are not representative for the majority in clinical practice and keeping them in the analyses may lead to clinically less relevant findings.[55,56] Trimming the extremes of the propensity score distributions is a data restriction strategy that generally will increase internal validity of findings.[57] 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. Matching in cohort studies does not require matched analyses, which simplifies the effect estimation to simple bivariate analyses. 1 : r matching allows consideration of all overlapping patients in the analysis but in variable ratio matching the matching sets need to be preserved in the analysis to avoid bias.



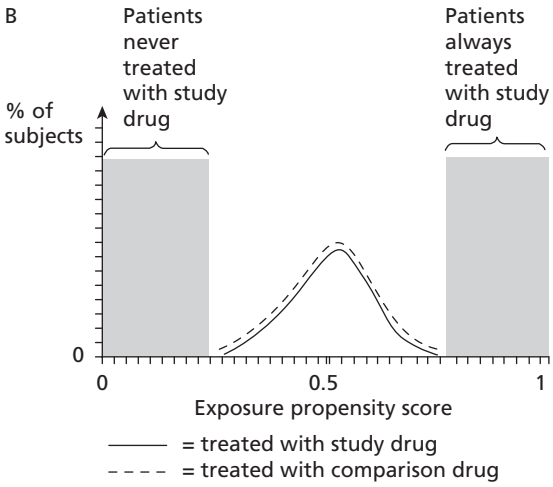


Figure 47.7 Two hypothetical propensity score distributions before and after matching: (A) Before matching: two propensity score distributions partially overlap indicating some similarities between the comparison groups in a multivariate parameter space. (B) After 1 : 1 matching on propensity score: not all patients found matches that were similar enough in their multivariable characteristics. Areas of non-overlap between propensity score distributions drop out entirely. Arrow, for each treated patient a person in the comparison group is identified with a similar propensity score. Reproduced from Schneeweiss[3] with permission from Wiley-Blackwell.

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.

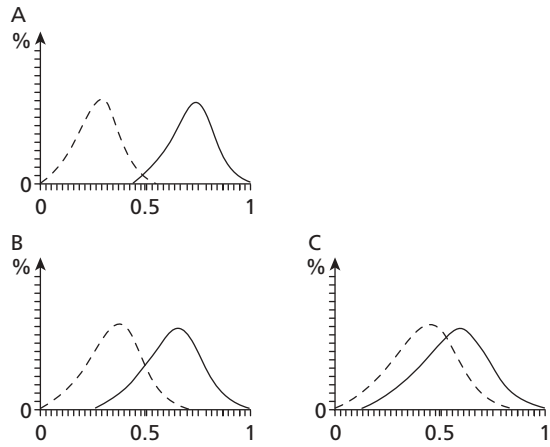


Figure 47.8 Multivariate propensity score distributions with varying degrees of overlap as a diagnostic tool.

It has been shown that on average multivariate covariate balance will be achieved between treatment groups when matching on propensity score.[58] If the 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 47.8A).[59] 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 a 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 a ezetimibe/ simvastatin combination continues to be marketed it will be used less selectively by more and more patients. Consequently, the propensity score distributions will overlap more and more as a sign that more patients are subject to treatment equipoise (Figure 47.8B and C).

If strong separation of PS distributions is observed, it means 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 health-care databases we can often define many covariates in an effort to reduce the limitation of unobserved or under-observed 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. 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[60] and either pragmatic pairwise matching to a common reference group or multidimensional matching is applied.[61] Of importance, PS analyses still can only adjust for measured variables, although they can be used to adjust for many at the same time. 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.[62,63] 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.[8] 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 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 a 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. 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 non-users or versus ongoing users of another drug.

Matching on patient characteristics. Multivariate propensity scores demonstrate areas of non-overlap 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[57] (see Figure 47.7B) 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.[9] 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 make choices accordingly.[9]

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 health-care databases, often lack critical details on health state and risk factors, which leads to residual confounding if left unadjusted.

Proxy adjustment

Longitudinal electronic health-care databases are as much a description of medical sociology under financial constraints as they are records of delivered health care and can be analyzed as a set of proxies that indirectly describe the health status of patients.[64] This status is presented through the lenses of health-care providers recording their findings and interventions via coders and operating under the constraints of a specific health-care system. On several steps along the way, weighting 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.[65] 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.[66,67] 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 health-care intensity are useful proxies for general health and access to care and have been shown to meaningfully help adjust for confounding.[68]

Proxy adjustment can be exploited by algorithms that systematically search through recorded codes for diagnoses, procedures, equipment purchases, and prescription drug dispensings before the initiation of study drug use to identify potential confounders or proxies thereof. The hundreds of proxies that will be identified can then be adjusted for in a large propensity score model. Collinearity may likely occur but is irrelevant, as the individual parameters estimated in the large propensity score regression will not be interpreted but only used for predicting treatment.[51] Such a high-dimensional propensity score approach has been empirically shown to improve confounding adjustment in many settings, although it is not yet fully evaluated. Although, adjusting for variables that are only related to the exposure and not to the outcome (an instrumental variable) could theoretically introduce bias,[53] 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.[69] The 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 above, 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, quasirandom treatment choices in routine care. Factors that determine such quasirandom 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 below).

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 based 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.[70] 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: (i) 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; (ii) an IV should be a factor that is as good as randomly assigned, so that it is unrelated to patient characteristics; and (iii) 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 non-compliance. The flip of a coin determines the instrument status (treat with A vs. treat with B) and the amount of random non-compliance determines the strength

of the instrument. In non-randomized 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.[71,72] Table 47.3 summarizes a list of some published IV analyses in health care.

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 non-selective NSAIDs (nsNSAID) on the risk of upper gastrointestinal (GI) bleed.[74] 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.[78] 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,[79,80] sicker patients may still cluster in coxib-preferring practices and be associated with GI bleed, which would invalidate the IV analysis. Stuckel *et al*.[75] 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 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:

$$\beta_{IV} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[X|Z=1] - E[X|Z=0]},$$

where Y is the study outcome and β is a measure of the effect of X on Y.[81] The numerator of this estimator is the effect of the instrument status (coxib-preferring physician vs. not) on the outcome

Table 47.3 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[71]
	Innovations and rapid adoption	Juurlink *et al.* Triamterene use in patients w/ hypertension before and after the RALES trial[72]
Provider treatment preference	Distance to specialist provider	McClellan *et al.* Distance to cardiac cath lab facility in patients with acute MI[73]
	Physician prescribing preference (PPP)	Brookhart *et al.* Physician's treatment initiation choice to the preceding patient[74]
	Regional treatment preference	Stukel *et al.* Variation if cardiac catheterization rates in 530 US regions in patients with MI[75]
	Hospital formulary/surgeon treatment preference	Schneeweiss *et al.* Cardiac surgeons who always use aprotinin as antifibrinolytic agent[70]
	Medication co-payment level	Cole *et al.* Medication copayment level in patients with heart failure and adherence[76]
	Dialysis center preference	Thamer *et al.* Epo dosing by non-profit vs. for-profit dialysis centers[77]

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 naïve risk difference estimate. As the instrument gets weaker, the denominator shrinks and the IV estimator increases relative to the naïve risk difference estimate. The denominator is sometimes called a rescaling parameter as it scales up the naïve 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 non-linear models for risk ratio estimation.[82] Brookhart *et al*. have suggested several empirical tests to investigate the quality of an instrument in health-care 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 are 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 health-care 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 above). 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.[83] The two-stage sampling approach samples patients according to their exposure and outcome status simultaneously and then reweights findings.[84] Collet *et al*. demonstrated two-stage sampling in a Canadian health-care database.[85]

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.[86] 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 now needs to be tied into the adjusted analysis of the main study cohort, as described below.

If additional information is available elsewhere, for example 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.[87,88] 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,[89] cognitive impairment, and physical impairments.[90] 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[88] 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).[91] 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.[92] 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.[93] 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 non-randomized 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 ($P_{C|E}$): [15]

$$RR_{adj} = \frac{RR_{unadj}}{\left[\frac{P_{C|E=1}(RR_{CD}-1)+1}{P_{C|E=0}(RR_{CD}-1)+1}\right]}$$

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.[94] 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 (www.drugepi.org).[95] Lash and Fink proposed an approach that considers several systematic errors simultaneously, allowing sensitivity analyses for confounding, misclassification, and selection bias in one process.[96]

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 6 months. In sensitivity analyses, this interval could be extended to 9 and 12 months. In a study on the comparative safety of antidepressant agents in children in British Columbia this interval was extended from 1 year to 3 years to ensure that the children in the study were treatment-naïve before their first use, which helped balance comparison groups and reduce confounding.[97] 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 tradeoff 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 health-care 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.

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 stronger such non-adherence 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 non-adherence, 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 non-adherence in an analysis of a drug effect requires information about the predictors of treatment discontinuation,[98,99] which is often not available with sufficient accuracy in pharmacoepidemiologic studies.

The future

Minimizing confounding in non-randomized pharmacoepidemiologic research is an ongoing development. While great progress has been made in analyzing longitudinal health-care 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 health-care 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, from disease registries, from patient surveys, and/or from 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**: 335–6.
2. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**: 915–20.
3. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010; **19**: 858–68.
4. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; **19**: 537–54.
5. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009; **20**: 488–95.
6. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol* 2008; **167**: 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, Brookhart MA, Avorn J, Maclure M, *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.
9. Psaty BM, Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. *JAMA* 2010; **304**: 897–8.
10. Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973; **29**: 479–86.
11. Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis appraisal by application to asbestos mining. *J R Stat Soc A* 1977; **140**: 469–91.
12. Lubin JH, Gail MH. Biased selection of controls for case-control analyses for cohort studies. *Biometrics* 1984; **40**: 63–75.
13. Spitzer WO, Suissa S, Ernst P, Horwitz RI, Habbick B, Cockcroft D, *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.
14. Breslow N, Day NE. *Statistical Methods in Cancer Research*. Volume I. The analysis of case-control studies. Lyon: International Agency for Research on Cancer, 1980.
15. Rothman KJ, Greenland S. *Modern Epidemiology*, 2nd edn. Hagerstown, MD: Lippincott-Raven, 1998.
16. Breslow N, Day NE. *Statistical Methods in Cancer Research*. Volume II. The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, 1987.
17. 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**: 87–92.
18. Suissa S, Hudson M, Ernst P. Leflunomide use and the risk of interstitial lung disease in rheumatoid arthritis. *Arthritis Rheum* 2006; **54**: 1435–9.
19. Suissa S, Edwardes MD, Boivin JF. External comparisons from nested case-control designs. *Epidemiology* 1998; **9**: 72–8.
20. 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**: 1749–57.
21. Suissa S, Dell'aniello S, Martinez C. The multitime case-control design for time-varying exposures. *Epidemiology* 2010; **21**: 876–83.
22. Hutchinson GB. Leukemia in patients with cancer of the cervix uteri treated with radiation. *J Natl Cancer Inst* 1968; **40**: 951–82.
23. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; **73**: 1–11.
24. 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**: 293–7.
25. 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**: 400–8.
26. Moulton LH, Wolff MC, Brenneman G, Santosham M. Case-cohort analysis of case-coverage studies of vaccine effectiveness. *Am J Epidemiol* 1995; **142**: 1000–6.
27. 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**: 2626–31.
28. Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* 1991; **2**: 155–8.
29. Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. *J Clin Epidemiol* 1999; **52**: 1165–72.
30. Wacholder S, Boivin JF. External comparisons with the case-cohort design. *Am J Epidemiol* 1987; **126**: 1198–209.
31. 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**: 27–31.
32. 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.
33. 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**: 845–9.
34. 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.
35. 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**: 991–7.
36. Meier CR, Derby LE, Jick SS, Vasilakis C, Jick H. Antibiotics and risk of subsequent first-time acute myocardial infarction. *JAMA* 1999; **281**: 427–31.

37. Barbone F, McMahon AD, Davey PG, Morris AD, Reid IC, McDevitt DG, *et al*. Association of road-traffic accidents with benzodiazepine use. *Lancet* 1998; **352**: 1331–6.
38. 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**: 319–26.
39. Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol* 1996; **143**: 1165–73.
40. Suissa S. The case-time-control design. *Epidemiology* 1995; **6**: 248–53.
41. Greenland S. Confounding and exposure trends in case-crossover and case-time-control design. *Epidemiology* 1996; **7**: 231–9.
42. Suissa S. The case-time-control design: further assumptions and conditions. *Epidemiology* 1998; **9**: 441–5.
43. 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**: 385–91.
44. 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**: 92–7.
45. Kjaer D, Horvath-Puho E, Christensen J, Vestergaard M, Czeizel AE, Sørensen HT, *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**: 181–8.
46. 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**: 159–67.
47. 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**: e1000157.
48. Risselada R, Straatman H, van Kooten F, Dippel DW, van der Lugt A, Niessen WJ, *et al*. Platelet aggregation inhibitors, vitamin K antagonists, and risk of subarachnoid haemorrhage. *J Thromb Haemost* 2011; **9**: 517–23.
49. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 2003; **12**: 551–8.
50. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005; **58**: 323–37.
51. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; **127**: 757–63.
52. 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): S114–20.
53. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006; **163**: 1149–56.
54. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol* 2006; **98**: 253–9.
55. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, *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**: 262–70.
56. Lunt M, Solomon D, Rothman K, Rothman K, Glynn R, Hyrich K, *et al*. Different methods of balancing covariates leading to different effect estimates in the presence of effect modification. *Am J Epidemiol* 2009; **169**: 909–17.
57. 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**: 843–54.
58. Rosenbaum PR, Rubin DB. The bias due to incomplete matching. *Biometrics* 1985; **41**: 103–16.
59. Sekula P, Caputo A, Dunant A, Roujeau JC, Mockenhaupt M, Sidoroff 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**: 10–18.
60. 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**: 2–9.
61. Solomon DH, Rassen JA, Glynn RJ, Garneau K, Levin R, Lee J, e*t al*. The comparative safety of opioids for

nonmalignant pain in older adults. *Arch Intern Med* 2010; **170**: 1979–86.
62. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res* 2009; **18**: 67–80.
63. Arbogast PG, Kaltenbach L, Ding H, Ray WA. Adjustment for multiple cardiovascular risk factors using a summary risk score. *Epidemiology* 2008; **19**: 30–7.
64. Schneeweiss S. Understanding secondary databases: a commentary on "Sources of bias for health state characteristics in secondary databases". *J Clin Epidemiol* 2007; **60**: 648–50.
65. Wolldridge JM. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2001.
66. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980; **112**: 564–9.
67. Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985; **122**: 495–506.
68. 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**: 854–64.
69. Myers JA, Rassen JA, Gagne JJ, Schneeweiss S, Huybrechts K, Rothman KJ, *et al*. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol* 2011, in press.
70. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med* 2008; **358**: 771–83.
71. 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**: 1539–56.
72. Juurlink DN, Mamdani MM, Lee DS, Kopp A, Austin PC, Laupacis A, *et al*. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med* 2004; **351**: 543–51.
73. 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**: 859–66.
74. 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**: 268–75.
75. 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**: 278–85.
76. Cole JA, Norman H, Weatherby LB, Walker AM. Drug copayment and adherence in chronic heart failure: effect on cost and outcomes. *Pharmacotherapy* 2006; **26**: 1157–64.
77. 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**: 1667–74.
78. 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**: 1233–41.
79. 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**: 3390–8.
80. 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 Supl 2): S116–22.
81. Angrist JD, Imbens GW, Rubin B. Identification of casual effects using instrumental variables. *J Am Stat Assoc* 1996; **94**: 444–55.
82. 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**: 273–84.
83. 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**: 297–305.
84. Walker AM. Anamorphic analysis: sampling and estimation for covariate effects when both exposure and disease are known. *Biometrics* 1982; **38**: 1025–32.
85. 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**: 309–15.

86. 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): S158–65.
87. Suissa S, Edwardes MD. Adjusted odds ratios for case-control studies with missing confounder data in controls. *Epidemiology* 1997; **8**: 275–80.
88. 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**: 17–24.
89. 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**: M750–6.
90. Eppig FJ, Chulis GS. Matching MCBS (Medicare Current Beneficiary Survey) and Medicare data: the best of both worlds. *Health Care Financ Rev* 1997; **18**: 211–29.
91. 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**: 279–89.
92. 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**: 734–45.
93. Sturmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Propensity score calibration and its alternatives. *Am J Epidemiol* 2007; **165**: 1122–3.
94. Psaty BM, Koepsell TD, Lin D, Weiss NS, Siscovick DS, Rosendaal FR, *et al*. Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc* 1999; **47**: 749–54.
95. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf* 2006; **15**: 291–303.
96. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003; **14**: 451–8.
97. Schneeweiss S, Patrick AR, Solomon DH, Dormuth CR, Miller M, Mehta J, *et al*. Comparative safety of antidepressant agents for children and adolescents regarding suicidal acts. *Pediatrics* 2010; **125**: 876–88.
98. 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**: 779–88.
99. 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**: 211–17.

PART VI
Conclusion

CHAPTER 48
The Future of Pharmacoepidemiology

Brian L Strom, Stephen E. Kimmel, and Sean Hennessy
Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, 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 is truly exploding, as is 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 1150 in 2011. The International Society for Pharmacoepidemiology (ISPE), established in 1991, has grown to over 1350 members from 45 countries. It has 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 2008.[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 its pharmacoepidemiologic manuscripts, as does 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 2.527 in 2009, similar to that of the *Journal of Clinical Epidemiology* and remarkably high for a niche field. The number of individuals seeking to enter the field continues to increase, as is 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 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 thereafter McGill University, Erasmus University Rotterdam, 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. 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

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

funding of external pharmacoepidemiologic studies. Requirements that a drug be shown to be cost-effective (see Chapter 38) have been added to many national health-care systems, provincial health-care systems, and managed care organizations, either to justify reimbursement or even to justify drug availability (see Chapter 31). Drug utilization review is being widely applied (see Chapter 25), and many hospitals are becoming mini-pharmacoepidemiologic practice and research laboratories. The US Congress has recognized the importance of pharmacoepidemiology, requiring FDA to build a new data resource, containing at least 100 million lives, for evaluating potential adverse effects of medical products (see Chapters 30 and 46).

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

The array of methodologic 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 Chapters 46 and 47). The next few years will likely see expanded use of propensity scores, instrumental variables, 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 36), especially given the controversies resulting from inconsistencies between non-experimental studies versus experimental studies, and given the emerging field of comparative effectiveness research (see Chapter 32).

Drug regulators have enthusiastically embraced therapeutic risk management (see Chapter 29). Yet, this field is 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 33). "Data mining" approaches will be used increasingly in spontaneous reporting databases to search for early signals of adverse reactions. 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 health-care claims or medical record data, is also clear (see Chapters 30 and 46).

We are likely to see increasing input from pharmacoepidemiologists into policy questions about drug approval (see Chapter 31). 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 examine patient- and regimen-specific factors that affect risk.[3] 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 Chapter 29).

With recent developments in molecular biology and bioinformatics, and their application to the

study of pharmacogenetics, exciting developments have occurred in the ability of researchers to identify biologic factors that predispose patients to adverse drug reactions[4] (see Chapter 34). 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, non-invasive 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. 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 automated databases, 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 and developed more. Studies of drug utilization will continue and will continue to become more innovative (see Chapter 24). Particularly as the health-care industry becomes more sensitive to the possibility of over-utilization, under-utilization, 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 25), potentially incorporating techniques from molecular pharmacoepidemiology (see Chapter 34). This is especially likely to be the case for studies of antibiotic misuse, as society becomes ever more concerned about the development of organisms resistant to currently available drugs (see Chapter 44).

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 use of computerized clinical decision support and the addition of pharmacists to patient care teams[12] (see Chapter 45).

Interest in the field of pharmacoeconomics, that is the application of the principles of health economics to the study of drug effects, is continuing (see Chapter 38). 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 non-experimental 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 37). 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 47). It is also being encouraged by the development of comparative effectiveness research (see Chapter 32). Other approaches to controlling for confounding are similarly likely to become more common as they are further developed (see Chapter 47).

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.

Research interest in the entire topic of patient non-adherence with prescribed drug regimens goes back to about 1960, but little fruitful research could be done until about 1990 because methods for ascertaining drug exposure in individual ambulatory patients were grossly unsatisfactory. The methodologic impasse was broken by two quite different developments. The initial one was to use very low doses of a very long half-life agent, phenobarbital, as a chemical marker, since a single measurement of phenobarbital in plasma is indicative of aggregate drug intake during the prior 2 weeks.[13] The other, more recent, advance has been to incorporate time-stamping microcircuitry into pharmaceutical containers, which records the date and time each time that 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 42) in research, and perhaps in clinical practice.

The next few years are also likely to see the increasing ability to target drug therapy to the proper patients. This will involve both increasing use of statistical methods, and increasing use of 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 will 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 health-care 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 increased 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 des 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 45), there has been increasing 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. 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 health care in particular, and the increased emphasis on using computerized 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 health maintenance organization [HMO] databases have emerged), and a number of new computerized databases have emerged as major resources for pharmacoepidemiologic research (e.g., commercial insurance databases [Chapter 13], inpatient databases [Chapter 16], the databases from Canada, Holland, and Denmark [Chapter 17 and 18]). 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 (Chapter 21) 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 14). 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 FDA's Sentinel Initiative[19] (see Chapter 30) will, similarly, provide a vast new data resource, eventually intended for hypothesis generating.

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 Chapters 22 and 23). 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. To address some of the problems inherent in any database, we must maintain the ability to perform *ad hoc* studies, as well. 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. Others will probably be developed as well.

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, as well, many different opportunities to form databases are being developed. There is also an 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] (see also Chapter 31).

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 FDA with a total of $1 million/year. Industry interest and support were similarly limited. With the increasing interest in the field, this situation appears to be changing rapidly. FDA is markedly expanding its intramural and extramural pharmacoepidemiologic program, and US National Institutes of Health (NIH) is increasingly funding pharmacoepidemiologic studies as well. Much more industry funding is available, as perceived need for the field within industry grows (see below). This is likely to increase, especially as the FDA expands its own pharmacoepidemiologic program, and more often requires industry to perform postmarketing studies.

There is, of course, a risk associated with academic groups becoming too dependent on industry funding, both in terms of choice of study questions and credibility. Fortunately, in the US, the Agency for Health Care Research and Quality (AHRQ) began 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 provides federal support for ongoing pharmacoepidemiologic activities (see also Chapter 6). While still small relative to industry expenditures on research, it is large relative to the US federal funding previously available for pharmacoepidemiology (see Chapter 32).

Even the US NIH has begun to fund pharmacoepidemiologic projects more often. NIH is the logical major US 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, NIH's current organizational structure represents an obstacle to pharmacoepidemiologic support. In general, the institutes within 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 the institute that 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 NIGMS for epidemiologic research. A notable exception is the NIGMS-funded Pharmacogenetics Research Network (PGRN), which has increasingly been performing larger scale pharmacogenetic epidemiologic studies. Further, NIGMS now funds one pharmacoepidemiologic training program, as part of its clinical pharmacologic training. 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.

Finally, but 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. As discussed in Chapter 35, this is already beginning to make pharmacoepidemiologic research more difficult, whether it is 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.

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 developed the Burroughs Wellcome Scholar Award in Pharmacoepidemiology, a faculty development award designed to bring new people into the field. This program, now discontinued, did not provide an opportunity for fellowship training of entry-level individuals, but was designed for more expe-

rienced investigators. Unfortunately, it is no longer an active program.

Outside of government, training opportunities are limited. In the US, the NIH is the major source of support for scientific training but as noted above, NIGMS, which funds training programs in clinical pharmacology, now supports one program in pharmacoepidemiology. The National Institute of Child Health and Human Development also funds training in pediatric pharmacoepidemiology. However, pharmacoepidemiologic training is still too dependent on non-federal 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 has been addressed, 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 32), additional training support is emerging from both NIH and AHRQ.

The view from industry

It appears that the role of pharmacoepidemiology in industry is and will continue to be 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 for industry 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 37) 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, FDA now has new authority to require postmarketing pharmacoepidemiologic studies, so one can expect to see many more required of industry by regulators.

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 described above. 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. Regulatory bodies 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 now plays an important role within the AHRQ, the Centers for Disease Control and Prevention, and the NIH, and there has been for 30 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 29) has been aggressively embraced by regulatory bodies around the world. This will continue to change regulation as more experience with it is gained.

Finally, there is an enormous increase in attention to drug safety, most recently driven by drug safety issues identified with COX-2 inhibitors and even traditional non-steroidal anti-inflammatory drugs, and then by the thiazolidinediones, used for treatment of diabetes. The net result has been major regulatory change, and even new legislation.

The view from the law

Finally, the importance of pharmacoepidemiology to the law has also been increasing. 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, therefore, 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 health-care 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. However, pharmacoepidemiology 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, Chen R, Dai WS, Dieck GS, *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. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**: 200–8.
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, Cole M, Verrill M, Jamieson D, *et al*. Influence of pharmacogenetics on response and toxicity in breast cancer patients treated with doxorubicin and cyclophosphamide. *Br J Cancer* 2010; **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, Lin PY, Tiamkao S, Khunarkornsiri U, *et al*. Strong association between HLA-B*5801 and allopurinol-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in a Thai population. *Pharmacogenet Genomics* 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, Laird N, Petersen LA, Teich JM, *et al*. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 1998; **280**: 1311–6.
13. Pullar T, Feely M. Problems of compliance with drug treatment: new solutions? *Pharm J* 1990; **245**: 213–5.
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?* National Academy 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. http://www.gao.gov/new.items/d10529t.pdf (accessed July 2011).
19. The Food and Drug Administration, eHealth Foundation, and the Brookings Institute. *Sentinel Initiative: Structure, Function and Scope*. Washington, D.C., December 16, 2008. http://www.fda.gov/oc/initiatives/criticalpath/transcript121608.pdf (accessed July 2011).
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 July 2011).
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: 11–7.
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. *1980 Report of the Joint Commission on Prescription Drug Use*. KF3869.R4. Rockville, MD.
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.

APPENDIX A
Sample Size Tables

Table A1. Sample sizes for cohort studies[a]

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	142 727	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	4 458
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	6 106	2 225
0.0005	39 401	55 751	126 078	588 332	756 333	210 078	63 015	32 669	20 999	15 117	11 662	7 870	4 219	2 845	1 215	439
0.001	19 694	27 865	63 015	294 037	377 953	104 973	31 483	16 320	10 488	7 549	5 823	3 928	2 104	1 418	603	216
0.005	3 928	5 557	12 564	58 600	75 249	20 888	6 257	3 240	2 080	1 495	1 152	775	412	276	114	37
0.01	1 957	2 769	6 257	29 170	37 411	10 378	3 104	1 605	1 028	738	568	381	201	133	53	15
0.05	381	538	1 212	5 627	7 140	1 969	582	297	188	133	101	65	32	19	4	—
0.10	184	259	582	2 684	3 357	918	266	133	82	57	42	26	10	4	—	—
0.15	118	166	372	1 703	2 095	568	161	79	47	32	23	13	—	—	—	—
0.20	85	120	266	1 212	1 465	393	109	52	30	19	13	6	—	—	—	—
0.25	65	92	203	918	1 086	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	—	—	—	—	—	—	—	—	—	—	—	—

[a] $\alpha = 0.05$ (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[a]

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	1529057	2153636	4825616	22279822	28149090	7764537	2302889	1183563	755529	540883	415381	278329	147626	99000	41938	15197
0.0001	152896	215349	482527	2227804	2814625	776367	230258	118337	151093	108167	83068	55659	29520	19795	8384	3036
0.0005	30570	43057	96475	445402	562673	155196	46024	23651	75539	54077	41528	27825	14756	9895	4189	1516
0.001	15280	21521	48218	222602	281179	77550	22994	11815	15095	10805	8297	5558	2946	1974	834	300
0.005	3047	4292	9613	44362	55984	15433	4571	2346	7540	5396	4143	2774	1469	984	414	148
0.01	1518	2138	4787	22082	27834	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	—	—	—	—	—	—	—	—	—	—	—	—

[a] $\alpha = 0.05$ (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[a]

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	1 369 471	1 930 847	4 322 614	19 888 657	24 913 372	6 843 626	2 014 756	1 029 014	653 418	465 696	356 275	237 254	124 571	83 030	34 793	12 510
0.00005	273 886	386 158	864 495	3 977 589	4 982 452	1 368 657	402 927	205 788	130 673	93 131	71 248	47 445	24 910	16 602	6 955	2 499
0.0001	136 938	193 072	432 230	1 988 706	2 491 087	684 286	201 449	102 885	65 330	46 560	35 619	23 719	12 452	8 299	3 476	1 248
0.0005	27 380	38 603	86 418	397 599	497 995	136 790	40 266	20 563	13 055	9 303	7 117	4 738	2 486	1 656	692	247
0.001	13 685	19 294	43 192	198 711	248 859	68 352	20 118	10 272	6 521	4 646	3 554	2 365	1 240	825	344	122
0.005	2 729	3 847	8 611	39 600	49 549	13 603	4 000	2 040	1 294	921	703	467	244	161	66	21
0.01	1 359	1 916	4 288	19 711	24 636	6 759	1 985	1 011	640	455	347	230	119	78	31	9
0.05	264	372	830	3 800	4 705	1 284	373	188	117	82	62	40	19	12	2	—
0.10	127	179	398	1 811	2 213	600	171	85	52	36	26	16	7	3	—	—
0.15	81	114	254	1 148	1 383	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	—	—	—	—	—	—	—	—	—	—	—	—

[a] $\alpha = 0.05$ (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[a]

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	—	—	—	—	—	—	—	—	—	—	—	—

[a] $\alpha = 0.05$ (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[a]

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	9 129	3 330
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	4 562	1 663
0.0005	29 433	41 645	94 178	439 474	564 968	156 925	47 071	24 404	15 686	11 292	8 712	5 879	3 152	2 126	908	329
0.001	14 711	20 816	47 071	219 641	282 325	78 413	23 518	12 191	7 835	5 639	4 350	2 935	1 572	1 060	451	162
0.005	2 935	4 152	9 385	43 774	56 210	15 604	4 675	2 421	1 554	1 117	861	579	309	207	86	28
0.01	1 463	2 069	4 675	21 790	27 946	7 752	2 319	1 199	769	552	425	285	151	100	40	12
0.05	285	402	906	4 204	5 334	1 471	435	222	141	100	76	49	24	15	3	—
0.10	138	194	435	2 005	2 508	686	200	100	62	43	32	20	8	4	—	—
0.15	89	125	278	1 273	1 566	425	121	59	36	24	17	10	—	—	—	—
0.20	64	90	200	906	1 095	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	—	—	—	—	—	—	—	—	—	—	—	—

[a] $\alpha = 0.05$ (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[a]

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	5 889	2 124
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	6 995	2 943	1 061
0.0005	23 799	33 257	73 580	335 708	417 346	114 455	33 648	17 182	10 914	7 783	5 958	3 973	2 091	1 396	586	210
0.001	11 895	16 622	36 775	167 779	208 557	57 193	16 812	8 584	5 452	3 887	2 975	1 983	1 043	696	292	104
0.005	2 372	3 315	7 332	33 436	41 526	11 382	3 343	1 705	1 082	771	589	392	205	136	56	19
0.01	1 182	1 651	3 651	16 643	20 647	5 656	1 659	845	536	381	291	193	100	66	26	8
0.05	230	321	707	3 208	3 944	1 075	312	157	99	69	52	34	17	10	2	—
0.10	111	154	339	1 529	1 856	503	144	71	44	30	23	14	6	3	—	—
0.15	71	99	216	969	1 160	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	—	—	—	—	—	—	—	—	—	—	—	—

[a] $\alpha = 0.05$ (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[a]

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	8 410
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	4 711	1 681
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	8 584	5 684	2 355	839
0.0005	21 759	30 314	66 590	301 015	368 057	100 225	29 111	14 721	9 276	6 570	5 001	3 305	1 714	1 134	469	166
0.001	10 875	15 151	33 281	150 439	183 927	50 082	14 545	7 354	4 634	3 282	2 498	1 650	855	566	233	82
0.005	2 169	3 021	6 635	29 979	36 623	9 968	2 892	1 461	920	651	495	326	168	111	45	15
0.01	1 080	1 505	3 304	14 922	18 210	4 954	1 436	725	456	322	245	161	83	54	21	6
0.05	210	292	639	2 876	3 480	942	271	135	84	59	44	29	14	8	2	—
0.10	101	140	306	1 370	1 638	441	125	62	38	26	19	12	5	2	—	—
0.15	65	90	195	868	1 025	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	—	—	—	—	—	—	—	—	—	—	—	—

[a] $\alpha = 0.05$ (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[a]

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	7 223
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	9 993	4 086	1 443
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	7 599	4 995	2 042	721
0.0005	20 685	28 795	63 057	283 636	343 387	93 087	26 824	13 473	8 442	5 950	4 510	2 960	1 518	997	407	143
0.001	10 338	14 392	31 515	141 754	171 599	46 516	13 402	6 731	4 217	2 972	2 253	1 478	757	497	203	71
0.005	2 061	2 870	6 282	28 248	34 169	9 259	2 665	1 338	837	590	446	292	149	98	39	13
0.01	1 027	1 429	3 128	14 059	16 990	4 601	1 323	663	415	292	221	144	73	48	19	6
0.05	199	277	605	2 709	3 247	876	250	124	77	53	40	26	12	8	2	—
0.10	96	133	289	1 290	1 529	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	—	—	—	—	—	—	—	—	—	—	—	—

[a] $\alpha = 0.05$ (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[a]

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	4 469
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	6 120	2 237
0.0005	39 412	55 772	126 151	588 806	757 227	210 362	63 120	32 734	21 046	15 155	11 695	7 896	4 238	2 862	1 228	451
0.001	19 704	27 887	63 088	294 510	378 847	105 257	31 588	16 384	10 535	7 587	5 856	3 954	2 124	1 435	617	228
0.005	3 939	5 579	12 638	59 074	76 145	21 173	6 363	3 304	2 127	1 533	1 184	801	432	293	128	49
0.01	1 968	2 790	6 331	29 646	38 309	10 663	3 210	1 669	1 076	777	601	407	221	150	67	27
0.05	391	560	1 288	6 111	8 059	2 261	690	363	237	172	135	93	52	37	18	9
0.10	195	281	659	3 181	4 302	1 219	379	202	133	98	77	54	32	23	13	8
0.15	129	189	451	2 215	3 072	879	278	150	100	75	60	43	26	19	11	8
0.20	97	143	348	1 741	2 476	716	230	126	85	64	52	37	23	18	11	8
0.25	77	116	287	1 465	2 137	624	203	113	77	59	48	35	23	18	12	9
0.30	64	98	248	1 289	1 930	569	188	106	73	56	46	34	23	18	13	10
0.35	56	86	222	1 174	1 802	536	180	103	72	56	46	35	24	19	14	11
0.40	49	77	203	1 097	1 727	519	177	102	72	56	47	36	25	20	15	12
0.45	44	70	191	1 048	1 694	513	178	104	74	58	49	38	27	22	17	14
0.50	40	66	182	1 023	1 696	519	182	108	77	61	52	40	29	24	19	16
0.55	38	62	178	1 019	1 732	535	191	114	82	66	56	44	32	27	21	18
0.60	36	61	177	1 035	1 806	562	203	123	89	72	61	49	36	31	25	21
0.65	35	60	180	1 077	1 927	605	222	135	99	80	69	56	42	36	29	25
0.70	34	61	188	1 149	2 110	669	248	153	113	92	79	64	49	43	35	31
0.75	35	64	203	1 268	2 390	764	287	178	133	109	94	77	59	52	43	38
0.80	37	70	230	1 465	2 831	913	348	218	164	135	117	97	75	66	55	49
0.85	43	82	278	1 811	3 591	1 168	451	285	216	179	156	129	101	90	75	68
0.90	54	108	379	2 527	5 143	1 687	659	420	320	266	233	195	154	137	116	105
0.95	93	190	690	4 717	9 851	3 257	1 288	828	635	531	466	391	313	280	238	217

[a] $\alpha = 0.05$ (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[a]

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	1529065	2153652	4825672	22280178	28149758	7764749	2302966	1183610	755564	540911	415405	278348	147639	99012	41948	15205
0.00005	305811	430731	965148	4456162	5630233	1553041	460628	236743	151128	108194	83091	55678	29534	19807	8393	3044
0.0001	152904	215366	482583	2228160	2815293	776578	230335	118385	75573	54105	41552	27844	14770	9906	4199	1524
0.0005	30578	43073	96531	445759	563340	155407	46101	23698	15130	10833	8321	5577	2960	1986	843	307
0.001	15288	21537	48274	222959	281846	77761	23072	11862	7574	5424	4167	2793	1483	996	424	155
0.005	3055	4308	9669	44719	56653	15644	4649	2393	1530	1097	844	567	302	204	88	34
0.01	1526	2154	4843	22440	28505	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

[a] $\alpha = 0.05$ (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[a]

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	1369478	1930861	4322663	19888975	24913964	6843813	2014824	1029056	653448	465720	356295	237271	124583	83040	34800	12517
0.00005	273893	386172	864545	3977907	4983044	1368844	402996	205830	130703	93155	71268	47461	24922	16612	6963	2506
0.0001	136945	193086	432280	1989023	2491679	684473	201517	102927	65360	46584	35640	23735	12464	8309	3483	1254
0.0005	27387	38617	86468	397917	498587	136977	40334	20604	13086	9328	7137	4754	2498	1666	700	253
0.001	13692	19309	43242	199028	249451	68540	20186	10314	6551	4671	3574	2382	1252	836	352	128
0.005	2736	3862	8661	39918	50143	13790	4068	2082	1324	945	724	484	256	171	73	28
0.01	1367	1931	4338	20030	25231	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

[a] $\alpha = 0.05$ (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[a]

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	1285573	1815890	4068256	18690963	23294197	6381647	1869301	950501	601245	427084	325786	215910	112440	74563	30952	11054
0.00005	257112	363178	813662	3738297	4659075	1276406	373889	190118	120262	85427	65166	43189	22493	14916	6193	2213
0.0001	128555	181589	406838	1869214	2329685	638251	186963	95070	60139	42720	32588	21599	11249	7461	3098	1108
0.0005	25709	36318	81379	373947	466173	127727	37422	19032	12041	8554	6526	4326	2255	1496	622	224
0.001	12853	18159	40697	187039	233234	63912	18729	9527	6028	4284	3269	2167	1130	750	313	113
0.005	2568	3632	8151	37513	46884	12860	3775	1923	1219	867	662	440	231	154	65	25
0.01	1283	1816	4082	18823	23592	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

[a] $\alpha = 0.05$ (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[a]

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	1472099	2082974	4710741	21983531	28264683	7851317	2355404	1221324	785139	565302	436191	294430	157960	106627	45676	16681
0.00005	294418	416596	942163	4396835	5653216	1570354	471115	244286	157043	113073	87248	58894	31598	21330	9139	3339
0.0001	147208	208299	471091	2198497	2826782	785234	235579	122156	78531	56544	43631	29452	15803	10668	4572	1671
0.0005	29440	41661	94233	439828	565636	157137	47150	24452	15721	11321	8736	5899	3166	2138	918	337
0.001	14719	20831	47126	219994	282992	78625	23596	12239	7870	5668	4375	2954	1587	1072	461	171
0.005	2943	4168	9441	44128	56879	15816	4753	2469	1589	1146	885	599	323	219	96	37
0.01	1470	2085	4730	22145	28617	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

[a] $\alpha = 0.05$ (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[a]

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	5 896	2 129
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	7 004	2 950	1 066
0.0005	23 805	33 269	73 622	335 976	417 842	114 612	33 705	17 216	10 939	7 803	5 975	3 986	2 101	1 405	593	216
0.001	11 901	16 635	36 817	168 047	209 054	57 349	16 869	8 618	5 477	3 907	2 993	1 997	1 053	705	298	109
0.005	2 378	3 327	7 374	33 704	42 024	11 540	3 400	1 740	1 107	791	607	406	215	145	63	24
0.01	1 188	1 664	3 693	16 911	21 146	5 814	1 717	880	561	402	308	207	211	75	33	14
0.05	236	333	750	3 482	4 455	1 237	371	193	125	91	70	48	27	19	10	5
0.10	117	167	383	1 810	2 383	669	205	109	71	53	41	29	17	13	7	5
0.15	77	112	261	1 258	1 704	484	152	82	54	41	32	23	14	11	7	5
0.20	58	84	201	987	1 376	396	126	69	47	35	28	21	13	10	7	5
0.25	46	68	166	829	1 190	346	112	62	43	33	27	20	13	10	7	5
0.30	38	57	143	729	1 076	316	105	59	41	32	26	20	13	11	7	6
0.35	33	50	127	662	1 006	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	1 016	318	116	70	52	42	36	29	22	19	15	13
0.65	20	34	101	601	1 085	343	127	78	58	47	40	33	25	22	18	16
0.70	20	34	105	640	1 190	380	143	89	66	54	47	38	29	26	21	19
0.75	20	35	112	705	1 350	435	166	104	78	64	56	46	36	32	26	23
0.80	21	38	126	812	1 601	520	201	127	96	80	70	58	45	40	34	30
0.85	23	44	152	1 002	2 034	667	261	167	127	106	93	77	61	55	46	42
0.90	29	58	205	1 395	2 916	965	383	246	189	158	139	117	94	84	71	65
0.95	48	100	371	2 598	5 592	1 868	750	487	376	316	279	236	190	171	147	134

[a] $\alpha = 0.05$ (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[a]

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	8 415
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	4 717	1 685
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	8 592	5 691	2 360	844
0.0005	21 764	30 325	66 628	301 253	368 496	100 363	29 161	14 751	9 298	6 588	5 016	3 316	1 723	1 141	474	171
0.001	10 881	15 162	33 319	150 678	184 367	50 220	14 595	7 384	4 655	3 299	2 513	1 662	864	573	239	87
0.005	2 174	3 032	6 672	30 218	37 064	10 107	2 943	1 491	942	668	510	338	177	118	50	19
0.01	1 086	1 516	3 342	15 161	18 652	5 093	1 486	755	478	340	259	173	91	61	27	11
0.05	215	303	678	3 120	3 932	1 085	323	167	107	77	60	41	23	16	8	4
0.10	107	152	345	1 620	2 105	588	179	94	62	45	35	25	15	11	6	4
0.15	70	101	235	1 125	1 507	426	132	71	47	35	28	20	12	9	6	4
0.20	52	76	181	882	1 218	349	111	60	41	31	25	18	11	9	6	4
0.25	42	62	149	741	1 053	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	1 061	339	128	80	60	49	42	35	27	23	19	17
0.75	17	31	99	624	1 204	389	149	93	70	58	51	42	32	29	24	21
0.80	18	33	111	718	1 429	466	181	115	87	72	63	52	41	37	31	28
0.85	20	38	132	884	1 817	598	235	151	115	96	84	70	56	50	42	38
0.90	25	50	179	1 230	2 607	867	345	223	172	144	127	107	85	77	65	59
0.95	41	85	323	2 288	5 002	1 678	678	442	342	288	255	215	174	157	134	123

[a] $\alpha = 0.05$ (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[a]

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	1034611	1440327	3154151	14188340	17179015	4657221	1342151	674222	422474	297830	225778	148193	76026	49982	20443	7227
0.00005	206920	288065	630838	2837745	3435981	931503	268452	134858	84505	59574	45162	29644	15209	9999	4091	1447
0.0001	103459	144032	315424	1418920	1718102	465788	134240	67438	42259	29792	22585	14825	7607	5002	2047	725
0.0005	20690	28806	63092	283861	343799	93216	26870	13501	8462	5966	4524	2970	1525	1003	412	147
0.001	10344	14403	31551	141978	172011	46645	13449	6759	4237	2988	2266	1489	765	504	207	75
0.005	2067	2880	6318	28473	34581	9388	2712	1366	858	606	460	303	157	104	44	17
0.01	1032	1440	3164	14285	17404	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

[a] $\alpha = 0.05$ (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

The *accuracy* of a measurement is the degree to which the measurement approximates the truth.

Ad hoc studies are studies that require primary data collection.

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 health-care 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.

An *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.

An *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.

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 ecological 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.

Anticipated harmful effects of drugs are unwanted effects that could have been predicted on the basis of existing knowledge.

An *association* is when two events occur together more often than one would expect by chance.

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

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.

Biological inference is the process of generalizing from a statement about an association seen in a population to a causal statement about biological 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.

An 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.*

Drug *clearance* is the proportion of the apparent volume of distribution that is cleared of 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, for example 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.

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.

A *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 an exposure and outcome.

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.

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.

Constructive knowledge, from a legal perspective, is knowledge that a person did not have, but could have acquired by the exercise of reasonable care.

A *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.

The *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 contrast 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).

The *direct medical costs* of medical care are the costs that are incurred in providing the care.

Direct non-medical costs are non-medical 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.

A *drug* is any exogenously administered substance that exerts a physiologic effect.

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 health-care system, whether national, regional, local, or institutional.

Ecological 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.

A 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.

A 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.

A study of drug *efficiency* is a study of whether a drug can bring about its desired effect at an acceptable cost.

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 well-being; several broad domains—physiological, functional, psychological, 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.

A *human research subject*, as defined in US regulation, 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." (Title 45 US Code of Federal Regulations Part 46.102 (f).)

Hypothesis-generating studies are studies that give rise to new questions about drug effects to be explored further in subsequent analytical 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.

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.

The *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 work force).

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.

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.

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.

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.

An *N-of-1 RCT* is a randomized controlled trial 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 didn't, 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.

Non-differential misclassification occurs when the misclassification of one variable does not vary by the level of another variable. Non-differential misclassification usually results in bias toward the null.

Non-experimental 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.

Observational studies (or non-experimental 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.

The *odds ratio* is the odds of exposure in the diseased group divided by the odds of exposure in the non-diseased 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.

An *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.

A *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.

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.

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.

A *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.

Pharmacovigilance is the identification and evaluation of drug safety signals. More recently, some have also used the term as synonymous with pharmacoepidemiology. 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" (WHO. Safety monitoring of medicinal products. *The Importance of Pharmacovigilance.* Geneva: World Health Organization, 2002). Mann defines *pharmacovigilance* as "the study of the safety of marketed drugs under the practical conditions of clinical usage in large communities" (Mann RD, Andrews EB, eds. *Pharmacovigilance*. Chichester: John Wiley & Sons Ltd, 2002).

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.

Pharmionics is the study of how patients use or misuse prescription drugs in ambulatory care.

Population-based databases or studies 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 determine incidence rates of diseases, as well as being more certain that one knows of all medical care that any given patient receives.

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 thought to apply to only 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 used 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 didn't, 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.

The *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).

Precision is the degree of absence of random error. Precise estimates have narrow confidence intervals.

Pre–post with comparison group design includes a single observation both before and after treatment in a non-randomly 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.

The *prevalence* of a disease is a measurement of how common the disease is. Specifically, it is the number of existing cases (both old and new 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, 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 medical 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, patient outcomes have not yet occurred as of the outset of the study.

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.

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 into 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, for example 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.

The *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*.

The *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.

A *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 upon repeated measurements of patients who have not changed.

Research, as defined in US regulation, is any activity designed to "develop or contribute to generalizable knowledge". (Title 45 US Code of Federal Regulations Part 46.102 (d).)

A *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." (US Code of Federal Regulations 46.102(f).)

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 as of the outset of the study.

Risk is the cumulative probability that something will happen.

A 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 a product's (drug or device) use.

Service registries are registries characterized by inclusion of subjects based on receipt of specific services, such as procedures, or based on hospitalizations.

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.

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.

A *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.

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.

Specific quality-of-life instruments are focused on disease or treatment issues specifically relevant to the question at hand.

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.

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.

A *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.

The *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 pharmacological 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.

A *type I statistical error* is concluding there is an association when in fact one does not exist, that is erroneously rejecting the null hypothesis.

A *type II statistical error* is concluding there is no association when in fact one does exist, that is 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.

An *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, for example from death (0.0) to full health (1.0). The key element of a utility instrument is that it is preference-based.

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.

Index

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