

Textbook of Pharmacoe**pidemiology**

Editors BRIAN L. STROM *and* STEPHEN E. KIMMEL



 **WILEY**

TEXTBOOK OF PHARMACOEPIDEMIOLOGY

TEXTBOOK OF PHARMACOEPIDEMOLOGY

Editors

BRIAN L. STROM and STEPHEN E. KIMMEL

University of Pennsylvania, Philadelphia, USA



John Wiley & Sons, Ltd

Copyright © 2006 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The Publisher is not associated with any product or vendor mentioned in this book.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, Canada L5R 4J3

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

Library of Congress Cataloging in Publication Data

Textbook of pharmacoepidemiology / edited by Brian L. Strom and Stephen E. Kimmel.

p. ; cm.

Includes bibliographical references and index.

ISBN-13: 978-0-470-02924-4 (hardback : alk. paper)

ISBN-10: 0-470-02924-2 (hardback : alk. paper)

ISBN-13: 978-0-470-02925-1 (pbk. : alk. paper)

ISBN-10: 0-470-02925-0 (pbk. : alk. paper)

1. Pharmacoepidemiology—Textbooks. I. Strom, Brian L. II. Kimmel, Stephen E.

[DNLM: 1. Pharmacoepidemiology—methods. 2. Drug Evaluation. 3. Drug Monitoring. 4. Drug Therapy.

5. Treatment Outcome. QZ 42 T3556 2006]

RM302.5.T49 2006

615'.7042—dc22

2006017302

British Library Cataloguing in Publication Data

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

ISBN-13 978-0-470-02924-4 (HB)

ISBN-10 0-470-02924-2 (HB)

ISBN-13 978-0-470-02925-1 (PB)

ISBN-10 0-470-02925-0 (PB)

Typeset in 9/11pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India

Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire

This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

List of Contributors	ix
Preface	xiii
Acknowledgments	xv
Acknowledgments from <i>Pharmacoepidemiology, Fourth Edition</i>	xvii
 SECTION I INTRODUCTION TO PHARMACOEPIDEMIOLOGY	 1
1. What is Pharmacoepidemiology? <i>Brian L. Strom</i>	3
2. Study Designs Available for Pharmacoepidemiology Studies <i>Brian L. Strom</i>	13
3. Sample Size Considerations for Pharmacoepidemiology Studies <i>Brian L. Strom</i>	25
4. Basic Principles of Clinical Pharmacology Relevant to Pharmacoepidemiology Studies <i>Sean Hennessy</i>	35
5. When Should One Perform Pharmacoepidemiology Studies? <i>Brian L. Strom</i>	55
6. Views from Academia, Industry, and Regulatory Agencies <i>Leanne K. Madre, Robert M. Califf, Robert F. Reynolds, Peter Arlett, and Jane Moseley</i>	63
 SECTION II SOURCES OF PHARMACOEPIDEMIOLOGY DATA	 89
7. Spontaneous Reporting in the United States <i>Syed Rizwanuddin Ahmad, Norman S. Marks, and Roger A. Goetsch</i>	91
8. Global Drug Surveillance: The WHO Programme for International Drug Monitoring <i>I. Ralph Edwards, Sten Olsson, Marie Lindquist, and Bruce Hugman</i>	117
9. Case–Control Surveillance <i>Lynn Rosenberg, Patricia F. Coogan, and Julie R. Palmer</i>	137
10. Prescription–Event Monitoring <i>Saad A.W. Shakir</i>	151
11. Overview of Automated Databases in Pharmacoepidemiology <i>Brian L. Strom</i>	167

12. Examples of Automated Databases	173
<i>Andy Stergachis, Kathleen W. Saunders, Robert L. Davis, Stephen E. Kimmel, Rita Schinnar, K. Arnold Chan, Deborah Shatin, Nigel S.B. Rawson, Sean Hennessy, Winanne Downey, MaryRose Stang, Patricia Beck, William Osei, Hubert G. Leufkens, Thomas M. MacDonald, and Joel M. Gelfand</i>	
13. Other Approaches to Pharmacoepidemiology Studies	215
<i>Brian L. Strom</i>	
14. How Should One Perform Pharmacoepidemiology Studies? Choosing Among the Available Alternatives	227
<i>Brian L. Strom</i>	
15. Validity of Pharmacoepidemiologic Drug and Diagnosis Data	239
<i>Suzanne L. West, Brian L. Strom, and Charles Poole</i>	
SECTION III SPECIAL ISSUES IN PHARMACOEPIDEMOLOGY METHODOLOGY	259
16. Bias and Confounding in Pharmacoepidemiology	261
<i>Ilona Csizmadi and Jean-Paul Collet</i>	
17. Determining Causation from Case Reports	277
<i>Judith K. Jones</i>	
18. Molecular Pharmacoepidemiology	287
<i>Stephen E. Kimmel</i>	
19. Bioethical Issues in Pharmacoepidemiologic Research	301
<i>Kevin Haynes, Jason Karlawish, and Elizabeth B. Andrews</i>	
20. The Use of Randomized Controlled Trials for Pharmacoepidemiology Studies	311
<i>Samuel M. Lesko and Allen A. Mitchell</i>	
21. The Use of Pharmacoepidemiology to Study Beneficial Drug Effects	321
<i>Brian L. Strom</i>	
22. Pharmacoeconomics: Economic Evaluation of Pharmaceuticals	333
<i>Kevin A. Schulman, Henry A. Glick, and Daniel Polsky</i>	
23. Using Quality-of-Life Measurements in Pharmacoepidemiologic Research	345
<i>Holger Schünemann, Gordon H. Guyatt, and Roman Jaeschke</i>	
24. The Use of Meta-analysis in Pharmacoepidemiology	353
<i>Carin J. Kim and Jesse A. Berlin</i>	
25. Patient Adherence to Prescribed Drug Dosing Regimens in Ambulatory Pharmacotherapy	367
<i>John Urquhart and Bernard Vrijens</i>	
26. Novel Approaches to Pharmacoepidemiology Study Design and Statistical Analysis	383
<i>Samy Suissa</i>	
SECTION IV SPECIAL APPLICATIONS OF PHARMACOEPIDEMOLOGY	397
27. Special Applications of Pharmacoepidemiology	399
<i>David Lee, Sumit R. Majumdar, Helene Levens Lipton, Stephen B. Soumerai, Sean Hennessy, Robert L. Davis, Robert T. Chen, Rosalie A. Bright, Allen A. Mitchell, David J. Graham, David W. Bates, and Brian L. Strom</i>	

28. The Future of Pharmacoepidemiology	447
<i>Brian L. Strom and Stephen E. Kimmel</i>	
Appendix A Sample Size Tables	455
Appendix B Glossary	473
Index	481

Contributors

SYED RIZWANUDDIN AHMAD, MD, MPH Medical Epidemiologist, Division of Drug Risk Evaluation, Office of Surveillance and Epidemiology, FDA, CDER, HFD-430, Room 3464, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA.

syed.ahmad@fda.hhs.gov

ELIZABETH B. ANDREWS, PhD, MPH Vice President, Pharmacoepidemiology and Risk Management, RTI Health Solutions, 3040 Cornwallis Road, PO Box 12194, Research Triangle Park, NC 27709-2194, USA.

eandrews@rti.org

PETER ARLETT, BSc, MBBS, MRCP, MFPM Principal Administrator, Pharmaceuticals Unit, European Commission, Honorary Senior Lecturer, Department of Medicine, University College London, 36 Fairbridge Road, London N19 3HZ, UK.

peter.arlett@ec.europa.eu

DAVID W. BATES, MD, MSc Chief, General Medicine Division, Brigham and Women's Hospital, 1620 Tremont Street, 3rd Fl, BC3-2M, Boston, MA 02120-1613, USA; Medical Director, Clinical and Quality Analysis, Partners HealthCare System, Inc.; Professor of Medicine, Harvard Medical School.

dbates@partners.org

PATRICIA BECK, BSP, MSc Research Consultant, Research Services, Saskatchewan Health, 3475 Albert Street, Regina, Saskatchewan S4S 6X6, Canada.

pbeck@health.gov.sk.ca

JESSE A. BERLIN, ScD Senior Director, Statistical Science, Johnson & Johnson Pharmaceutical Research and Development, LLC, 1125 Trenton-Harbourton Road, PO Box 200 (mail stop 67), Titusville, NJ 08560, USA.

jberlin@prdus.jnj.com

ROSELIE A. BRIGHT, ScD Epidemiologist, Center for Devices and Radiological Health, United States Food and Drug Administration, 1350 Piccard Drive, HFZ-541, Rockville, MD 20850, USA.

roselie.bright@fda.hhs.gov

ROBERT M. CALIFF, MD Vice Chancellor for Clinical Research, Director, Duke Clinical Research Institute, Professor of Medicine, Duke University Medical Center, 2400 Pratt Street, Room 0311 Terrace Level, Durham, NC 27705, USA.

robert.califf@duke.edu

K. ARNOLD CHAN, MD, ScD Associate Professor of Medicine, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA.

kachan@hsph.harvard.edu

ROBERT T. CHEN, MD, MA Injection Safety Coordinator, Global Aids Program, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30333, USA.

rtc1@cdc.gov

JEAN-PAUL COLLET, MD, PhD, MSc Professor, Department of Pediatrics, UBC Associate Director, Partnership Development, CFRI K-103, Ambulatory Care Building, 4480 Oak Street, Vancouver, BC, Canada V6H 3V4.

JCollet@cw.bc.ca

PATRICIA F. COOGAN, ScD Associate Professor of Epidemiology, Slone Epidemiology Center, Boston University, 1010 Commonwealth Avenue, 4th floor, Boston, MA 02215, USA.

pcoogan@slone.bu.edu

ILONA CSIZMADI, PhD, MSc Research Scientist/Epidemiologist, Division of Population Health and Information, Alberta Cancer Board, Adjunct Assistant Professor, Department of Community Health Sciences

Faculty of Medicine, University of Calgary, 1331-29 Street NW, Calgary, Alberta T2N 4N2, Canada.
ilona.csizmadi@cancerboard.ab.ca

ROBERT L. DAVIS, MD, MPH Director, Immunization Safety Office, Office of the Chief Science Officer, Centers for Disease Control and Prevention, 1600 Clifton Road NE MS D-26 Atlanta, GA 30333, USA.
rad2@cdc.gov

WINANNE DOWNEY, BSP Manager, Research Services, Saskatchewan Health, 3475 Albert Street, Regina, Saskatchewan S4S 6X6, Canada.
wdowney@health.gov.sk.ca

I. RALPH EDWARDS, MB, ChB Professor and Director, WHO Collaborating Centre for International Drug Monitoring, Uppsala Monitoring Centre, Stora Torget 3, S-753 20 Uppsala, Sweden.
ralph.edwards@who-umc.org

JOEL M. GELFAND, MD, MSCE Medical Director, Clinical Studies Unit, Assistant Professor of Dermatology, Associate Scholar, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, 3600 Spruce Street, 2 Maloney Building, Philadelphia, PA 19104, USA.
joel.gelfand@uphs.upenn.edu

HENRY A. GLICK, PhD Assistant Professor of Medicine, Division of General Internal Medicine, University of Pennsylvania School of Medicine, 1211 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, USA.
hlthsvrs@mail.med.upenn.edu

ROGER A. GOETSCH, RPh, PharmD Special Assistant for Regulatory Affairs, Electronic Submission Coordinator, Office of Surveillance and Epidemiology (HFD-430), 12300 Twinbrook Parkway, Suite 240, Rockville, MD 20851, USA.
roger.goetsch@fda.hhs.gov

DAVID J. GRAHAM, MD, MPH Associate Director for Science and Medicine, Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, 10903 New Hampshire Avenue, Building 22, Room 4314, Silver Spring, MD 20993, USA.
david.graham1.fda.hhs.gov

GORDON H. GUYATT, MD Professor, Clinical Epidemiology and Biostatistics, Department of Medicine, McMaster University, Hamilton, Ontario L8N 3Z5, Canada.
guyatt@mcmaster.ca

KEVIN HAYNES, PharmD Pharmacoepidemiology Postdoctoral Fellow, Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 108 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021 USA.
khaynes@cceb.med.upenn.edu

SEAN HENNESSY, PharmD, PhD Assistant Professor of Epidemiology and Pharmacology, Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 803 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021 USA.
shenness@cceb.med.upenn.edu

BRUCE HUGMAN, BA, MA, Diploma in Education Communications Consultant to the Uppsala Monitoring Centre, Senior Academic Adviser, Naresuan University, Phayao Campus, Thailand, PO Box 246, Amphur Muang, Chiang Rai 57000, Thailand.
mail@brucehugman.net

ROMAN JAESCHKE, MD, MSc Clinical Professor of Medicine, Department of Medicine, St. Joseph's Hospital, 301 James Street S, Hamilton, Ontario L8N 3A6, Canada.
jaeschke@mcmaster.ca

JUDITH K. JONES, MD, PhD President, The Degge Group, Ltd., Suite 1430, 1616 N. Ft. Myer Drive, Arlington, VA 22209-3109, USA.
jkjones@deggegroup.com

JASON KARLAWISH, MD Associate Professor of Medicine, Institute on Aging, Division of Geriatrics and Center for Bioethics, University of Pennsylvania School of Medicine, 3615 Chestnut Street, Philadelphia, PA 19104, USA.
jasonkar@mail.med.upenn.edu

CARIN J. KIM, MS Graduate Student, Division of Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, 501 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, USA.
cakim@cceb.upenn.edu

STEPHEN E. KIMMEL, MD, MSCE Associate Professor of Medicine and Epidemiology, Department of Medicine, Cardiovascular Division, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 717 Blockley Hall, 423 Guardian

Drive, Philadelphia, PA 19104, USA.
skimmel@cceb.med.upenn.edu

DAVID LEE, MD Deputy Director, Technical Strategy and Quality, Center for Pharmaceutical Management, Management Sciences for Health, Inc., 4301 N. Fairfax Drive, Suite 400, Arlington, VA 22203-1627, USA.
dlee@msh.org

SAMUEL M. LESKO, MD, MPH Director of Research and Medical Director, Northeast Regional Cancer Institute, University of Scranton Campus, 334 Jefferson Avenue, Scranton, PA 18510, USA.
leskos2@scranton.edu

HUBERT G. LEUFKENS, PhD Dean of Pharmaceutical Sciences, Utrecht University, Faculty of Science, Utrecht Institute for Pharmaceutical Sciences, Division of Pharmacoepidemiology and Pharmacotherapy, PO Box 80082, 3508 TB, Utrecht, The Netherlands.
H.G.M.Leufkens@pharm.uu.nl

MARIE LINDQUIST, MSc, PhD Head, Data Management and Research, General Manager, Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, Uppsala Monitoring Centre, Stora Torget 3, S-753 20 Uppsala, Sweden.
marie.lindquist@who-umc.org

HELENE LEVENS LIPTON, PhD Professor of Health Policy, Department of Clinical Pharmacy, School of Pharmacy, Institute for Health Policy Studies, School of Medicine, University of California at San Francisco, 3333 California Street, Suite 420, St. Laurel Heights, San Francisco, CA 94118, USA.
liptonh@pharmacy.ucsf.edu

THOMAS M. MACDONALD MD, FRCP, FESC, FISPE, FBPharmacolS Professor of Clinical Pharmacology and Pharmacoepidemiology, Department of Clinical Pharmacology and Therapeutics, Division of Medicine and Therapeutics, University of Dundee, Ninewells Hospital and Medical School, Dundee, Scotland, DD1 9SY, UK.
t.m.macdonald@dundee.ac.uk

LEANNE K. MADRE, JD, MHA Program Director, CERTs Coordinating Center, Duke University Medical Center, PO Box 17969, Durham, NC 27715, USA.
leanne.madre@duke.edu

SUMIT R. MAJUMDAR, MD, MPH Associate Professor, Division of General Internal Medicine, Department of

Medicine, University of Alberta, Edmonton, Alberta T6G 2B7, Canada.
me2.majumdar@ualberta.ca

NORMAN S. MARKS, MD, MHA Medical Director, MedWatch Program, US Food and Drug Administration, CDER/OCD, Safety Policy and Communication Staff, Rockwall II, Suite 5100, 5515 Security Lane, Rockville, MD 20852, USA.
norman.marks@fda.hhs.gov

ALLEN A. MITCHELL, MD Director, Slone Epidemiology Center, Professor of Epidemiology and Pediatrics, Boston University Schools of Public Health & Medicine, 1010 Commonwealth Avenue, Boston, MA 02215, USA.
amitchell@slone.bu.edu

JANE MOSELEY, MSc, MFPM Team Leader, Pharmacoepidemiology Research Team, Medicines and Healthcare Products Regulatory Agency (MHRA), Room 15-206, Market Towers, 1 Nine Elms Lane, London SW8 5NQ, UK.
jane.moseley@mhra.gsi.gov.uk

STEN OLSSON, MSc, Pharm Head of External Affairs, Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, Uppsala Monitoring Centre, Stora Torget 3, S-753 20 Uppsala, Sweden.
sten.olsson@who-umc.org

WILLIAM OSEI, MD, MPH Provincial Epidemiologist, Saskatchewan Health, 3475 Albert Street, Regina, Saskatchewan S4S 6X6, Canada.
wosei@health.gov.sk.ca

JULIE R. PALMER, ScD Professor of Epidemiology, Boston University School of Public Health, Senior Epidemiologist, Slone Epidemiology Center at Boston University, 1010 Commonwealth Avenue, Boston, MA 02215, USA.
jpalmer@slone.bu.edu

DANIEL POLSKY, PhD Research Associate Professor of Medicine, Division of General Internal Medicine, University of Pennsylvania School of Medicine, 1212 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, USA.
polsky@mail.med.upenn.edu

CHARLES POOLE, MPH, ScD Associate Professor, Department of Epidemiology (CB 7435), University of North Carolina School of Public Health, Pittsboro Road, Chapel Hill, NC 27599-7435, USA.
poolec@email.unc.edu

NIGEL S.B. RAWSON, MSc, PhD Pharmacoepidemiologist, GlaxoSmithKline, 7333 Mississauga Road, Mississauga, Ontario, 5N 6L4, Canada.
nigel.s.rawson@gsk.com

ROBERT F. REYNOLDS, ScD Senior Director, Head, Global Epidemiology, Pfizer, Inc., 235 East 42nd Street, New York, NY 10017, USA.
robert.reynolds@pfizer.com

LYNN ROSENBERG, ScD Associate Director, Slone Epidemiology Center, Boston University, 1010 Commonwealth Ave, 4th Floor, Boston, MA 02215, USA.
lrosenberg@slone.bu.edu

KATHLEEN W. SAUNDERS, JD Analyst/Programmer, Center for Health Studies, Group Health Cooperative, 1730 Minor Ave, Suite 1600, Seattle, WA 98101, USA.
saunders.k@ghc.org

RITA SCHINNAR, MPA Senior Research Analyst and Project Manager, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 807 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, USA.
rschinnar@cceb.med.upenn.edu

KEVIN A. SCHULMAN, MD Professor of Medicine and Management, Director, Center for Clinical and Genetic Economics, Duke Clinical Research Institute, Duke University Medical Center, PO Box 17969, Durham, NC 27715, USA.
kevin.schulman@duke.edu

HOLGER SCHÜNEMANN, MD, PhD, FACP, FCCP Associate Professor of Medicine and Preventive Medicine, Clinical Epidemiology and Biostatistics, Department of Epidemiology, Italian National Cancer Institute Regina Elena, Rome Via Elio Chianesi 53, 00144 Rome, Italy.
hjs@buffalo.edu or schuneh@mcmaster.ca

SAAD A.W. SHAKIR, MB, ChB, LRCP&S, FRCP, FFPM, MRCGP, FISPE Director, Drug Safety Research Unit, Bursledon Hall, Blundell Lane, Southampton SO31 1AA, UK.
saad.shakir@dsru.org

DEBORAH SHATIN, PhD Shatin Associates, 10030 30th Avenue North, Plymouth, MN 55441, USA.
dshatin04@aol.com

STEPHEN B. SOUMERAI, ScD Professor of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 133 Brookline Avenue, 6th Floor, Boston, MA 02215, USA.
ssoumerai@hms.harvard.edu

MARYROSE STANG, PhD Research Consultant, Research Services, Saskatchewan Health, 3475 Albert Street, Regina, Saskatchewan S4S 6X6, Canada.
mstang@health.gov.sk.ca

ANDY STERGACHIS, PhD, RPh Professor of Epidemiology and Adjunct Professor of Pharmacy, Northwest Center for Public Health Practice, School of Public Health & Community Medicine, University of Washington, 1107 NE 45th Street, Ste 400, Seattle, WA 98105, USA.
stergach@u.washington.edu

BRIAN L. STROM, MD, MPH George S. Pepper Professor of Public Health and Preventive Medicine, Professor of Biostatistics and Epidemiology, Medicine, and Pharmacology, Chair, Department of Biostatistics and Epidemiology, Director, Center for Clinical Epidemiology and Biostatistics, Associate Vice Dean, University of Pennsylvania School of Medicine, Associate Vice President for Strategic Integration, University of Pennsylvania Health System, 824 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, USA.
bstrom@cceb.med.upenn.edu

SAMY SUISSA, PhD James McGill Professor of Epidemiology and Biostatistics and of Medicine, McGill University, Director, Division of Clinical Epidemiology, Royal Victoria Hospital, 687 Pine Avenue West, R4.29, Montreal, Quebec H3A 1A1, Canada.
samy.suissa@clinepi.mcgill.ca

JOHN URQUHART, MD 975 Hamilton Ave, Palo Alto, CA 94301, USA.
urquhart@ix.netcom.com

BERNARD VRIJENS, PhD 24 Rue des Cyclistes Frontière, BE 4600 Visé, Belgium.
bernard.vrijens@pharmionics.eu

SUZANNE L. WEST, MPH, PhD Research Associate Professor, Department of Epidemiology, Cecil G. Sheps Center for Health Services Research, 725 Martin Luther King Boulevard, CB 7521, University of North Carolina, Chapel Hill, NC 27599-7521, USA.
sue_west@unc.edu

Preface

It was a remarkable 17 years ago that the first edition of Strom's *Pharmacoepidemiology* was published. The preface to that book stated that pharmacoepidemiology was a new field with a new generation of pharmacoepidemiologists arising to join the field's few pioneers. Over the ensuing 17 years, the field indeed has grown and no longer deserves to be called "new." Many of that "new generation" (including one of the Editors of this book) are now "middle-aged" pharmacoepidemiologists. Despite its relatively brief academic life, a short history of pharmacoepidemiology and a review of its current state will set the stage for the purpose of this Textbook.

Pharmacoepidemiology originally arose from the union of the fields of Clinical Pharmacology and Epidemiology. Pharmacoepidemiology studies the use of and the effects of drugs in large numbers of people and applies the methods of Epidemiology to the content area of Clinical Pharmacology. This field represents the science underlying post-marketing 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 first publication of Strom. The International Society of Pharmacoepidemiology has grown into a major international scientific force, with over 800 members from 49 countries, an extremely successful and well-attended annual meeting, a large number of very active committees, and its own journal. 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 vigorous Pharmacoepidemiology Section within the American Society for Clinical Pharmacology and Therapeutics, and with Pharmacoepidemiology a required part of the Clinical Pharmacology board examination.

Most of the major international pharmaceutical companies have founded special units to organize and lead their efforts in pharmacoepidemiology, pharmacoeconomics, and quality-of-life studies. The seeming explosion of drug safety crises continues to emphasize the need for the field, and some foresighted manufacturers have begun to perform "prophylactic" pharmacoepidemiology studies, in order 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. The United States has recently implemented a new drug benefit as part of

Medicare—paying for drugs for those over age 65. This will generate an enormous new interest in this field, and potentially new data and new funding, as our federal government becomes concerned about the costs and effects of prescription drugs. 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 summary, there has been tremendous growth in the field of Pharmacoepidemiology and a fair amount of maturation. With the growth and maturation of the field, Strom's *Pharmacoepidemiology* has grown and matured along with it. As a reflection of the growth of the field, the fourth edition of Strom is over twice as long as the first! *Pharmacoepidemiology* thus represents a comprehensive source of information about the field.

So why, one may ask, do we need a *Textbook of Pharmacoepidemiology*? The need arose precisely because of the growth of the field. With that, and the corresponding growth in the parent book, Strom's *Pharmacoepidemiology* has really become more of a reference book than a book usable as a textbook. Yet, there is increasing need for people to be trained in the field and for an increasing number of training programs. With the maturity of the field, therefore, comes the necessity for both comprehensive approaches (such as Strom's *Pharmacoepidemiology*) and more focused approaches. As such, *Textbook of Pharmacoepidemiology* is intended as a modified and shortened version of its parent, designed to meet the need of students. We believe that students can benefit from an approach that focuses on the core of the discipline, along with learning aids.

Textbook of Pharmacoepidemiology attempts to fill this need, providing a focused educational resource for students. It is our hope that this book will serve as a useful textbook for students at all levels: upper-level undergraduates, graduate students, post-doctoral fellows, and others who are learning the field. In order to achieve our goals, we have substantially shortened Strom's *Pharmacoepidemiology* with a focus on what is needed by students, eliminating some chapters and shortening others. We also have provided Case Examples for most chapters and Key Points for all chapters. Each chapter is followed by a list of Suggested Further Readings.

Specifically, we have tried to emphasize the methods of Pharmacoepidemiology and the strengths and limitations of

the field, while minimizing some of the technical specifications that are important for a reference book but not for students. As such, the first five chapters of Section I, "Introduction to Pharmacoepidemiology," lay out the cores of the discipline, and remain essentially unchanged from Strom's *Pharmacoepidemiology*, with the exception of the inclusion of Key Points and lists of Suggested Further Readings. We have also included a chapter on different perspectives of the field (Chapter 6), as a shortened form of several chapters from the original. Section II focuses on Sources of Pharmacoepidemiology Data and includes important chapters about spontaneous reporting, methods of surveillance for adverse drug reactions, and other approaches to pharmacoepidemiology studies. A substantially shortened chapter on Examples of Automated Databases (Chapter 12) is included, focused on the strengths and limitations of these data sources rather than providing extensive details about the content of each database. Critical to understanding the strengths and limitations of using any data is the validity of those data. The chapter on Validity of Pharmacoepidemiologic Drug and Diagnosis Data (Chapter 15) is thus a key part of Section II. Section III summarizes Special Issues in Pharmacoepidemiology Methodology that we feel would be important to more advanced pharmacoepidemiology students. Although no student is likely to become an expert in all of these methodologies, they form a core set of knowledge that we believe all pharmacoepidemiologists should have. In addition, one never knows what one will do later in one's own career, nor when one may be called upon to help others with the use of these methodologies. Section IV concludes the textbook with a collection of Special Applications of the field, and speculation about its future, always an important consideration for young investigators when charting a career path.

Pharmacoepidemiology may be maturing, but many exciting opportunities and challenges lie ahead as the field continues to grow and respond to unforeseeable future events. It is our hope that this book can serve as a useful introduction and resource for students of Pharmacoepidemiology, both those enrolled in formal classes and those learning in "the real world," who will respond to the challenges that lie ahead. Of course, we are always students of our own discipline, and the process of developing this Textbook has been educational for us. We hope that this book will also be stimulating and educational for you.

Brian L. Strom, MD, MPH, 2006
Stephen E. Kimmel, MD, MSCE, 2006

Acknowledgments

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 by numerous grants from government, foundations, and industry. 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. We would also like to thank our publisher, John Wiley & Sons, Ltd, for their assistance and insights in support of this book.

Rita Schinnar and Edmund Weisberg were instrumental in helping to shorten several of the chapters that were merged together from Strom's *Pharmacoepidemiology*, and Rita also provided additional editorial assistance with many other chapters. Anne Saint John did an amazing job of formatting and organizing the myriad versions of all the book's chapters. We also would like gratefully to acknowledge the authors of those chapters in *Pharmacoepidemiology Fourth Edition* from which material in this book was adapted. A full listing of these authors appears separately overleaf. Finally, we would like to thank all of the authors for the work that they did in helping to revise their chapters and provide case examples, key points, and suggested readings to this textbook.

We would like to thank our parents for the support and education that formed the basis of our careers. B.L.S. would also like to thank Paul D. Stolley, MD, MPH and the late Kenneth L. Melmon, MD for their direction, guidance, and inspiration in the formative years of his career. S.E.K. expresses his eternal gratitude to his mentor, B.L.S., for 15 years of mentoring, guidance, and support, and for the chance to work on this book. We would also like to thank our trainees, from whom we have learned so much over the years and who were the inspiration for this book. Relatedly, B.L.S. would like to thank S.E.K. for joining him in this effort for the first time, taking on a disproportionate share of the new work. Not only was it a true pleasure to work together, as always, but as a former trainee, in this effort to create a resource for new trainees, it made me (B.L.S.) particularly proud to work with you (S.E.K.).

This book would not have been possible without the love, understanding, and support of our families: The Stroms—Lani, Shayna, and Jordi; and The Kimmels—Alison, David, Benjamin, and Jonathan. We have the double fortune of having the support of not only our own families but the support of each other's family as well.

Acknowledgments from *Pharmacoepidemiology, Fourth Edition*

The following authors contributed to those chapters in *Pharmacoepidemiology, Fourth Edition* that have been adapted in creating this book.

Mark I. Avigan
Ulf Bergman
Jean-Francois Boivin
Arthur Caplan
Jeffrey L. Carson
David Casarett
Hassy Dattani
Gretchen S. Dieck
Gary D. Friedman
Kate Gelperin
Dale B. Glasser
Margaret J. Gunter
Jerry H. Gurwitz
David A. Henry
Lisa J. Herrinton
Eric S. Johnson
Rainu Kaushal
David J. Margolis

Bentson H. McFarland
Patricia McGettigan
Kenneth L. Melmon
Andrew Mosholder
Winnie W. Nelson
James L. Nichol
John Parkinson
Richard Platt
Marsha A. Raebel
Wayne A. Ray
Timothy Rebbeck
Philip H. Rhodes
Douglas W. Roblin
Joe V. Selby
Paul J. Seligman
David H. Smith
Anne Tonkin
Li Wei

SECTION I

INTRODUCTION TO PHARMACOEPIDEMIOLOGY

1

What is Pharmacoepidemiology?

Edited by:

BRIAN L. STROM

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 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. Recent data suggest that perhaps 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. 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 PHARMACOEPIDEMOLOGY

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 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.

PHARMACOEPIDEMOLOGY 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 4). 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 the infection, it is not sufficient to know that gentamicin has a small probability of causing renal disease. A good clinician should realize that a patient who has impaired liver function is at a greater risk of suffering from this adverse effect than one with normal liver function. 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, 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 unusual application of the field.

Specifically, the field of pharmacoepidemiology has primarily concerned itself with the study of adverse drug effects. Adverse reactions have traditionally been separated into those which are the result of an exaggerated but otherwise usual pharmacological effect of the drug, sometimes

called *Type A reactions*, versus those which are aberrant effects, so-called *Type B reactions*. 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. Third, they may have normal drug levels, but for some reason are overly sensitive to them.

In contrast, Type B reactions tend to be uncommon, not related to dose, unpredictable, and potentially more serious. They usually require cessation of the drug. They may be due to what are known as hypersensitivity reactions or immunologic reactions. Alternatively, Type B reactions may be some other idiosyncratic reaction to the drug, either due to some inherited susceptibility (e.g., glucose-6-phosphate dehydrogenase deficiency) or due to some other mechanism. Regardless, Type B reactions are the more difficult to predict or even detect, and represent the major focus of many pharmacoepidemiology studies of adverse drug reactions.

The usual approach to studying adverse drug reactions has been the collection of spontaneous reports of drug-related morbidity or mortality (see Chapters 7 and 8). However, determining causation in case reports of adverse reactions can be problematic (see Chapter 17), as can attempts to compare the effects of drugs in the same class. This has led academic investigators, industry, the Food and Drug Administration (FDA), and the legal community to turn to the field of epidemiology. Specifically, *studies of adverse effects* have been supplemented with *studies of adverse events*. In the former, investigators examine case reports of purported adverse drug reactions and attempt to make a subjective clinical judgment on an *individual* basis about whether the adverse outcome was actually caused by the antecedent drug exposure. In the latter, controlled studies are performed examining whether the adverse outcome under study occurs more often in an exposed *population* than in an unexposed population. This marriage of the fields of clinical pharmacology and epidemiology has resulted in the development of a new field: pharmacoepidemiology.

PHARMACOEPIDEMOLOGY VERSUS EPIDEMIOLOGY

Epidemiology is the study of the distribution and determinants of diseases in populations (see Chapter 2). 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, i.e., epidemics. More recently, it has also been concerned with 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 (see Chapter 20), 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.

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

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. In response, the Food, Drug, and Cosmetic Act was passed in 1938. 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. In 1952, the first textbook of adverse drug reactions was published. 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. 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 (see Chapter 27). This approach was later to be transported to the University of Florida–Shands Teaching Hospital as well.

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. 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 8).

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 pharmacological and toxicological testing before a drug could be tested in humans. The data from these studies were required to be submitted to the FDA in an Investigational New Drug Application (IND) 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 Drug Efficacy Study Implementation (DESI) process, conducted by the National Academy of Sciences' National Research Council with support from a contract from the FDA, was not completed until relatively recently, 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. However, the drugs that are marketed are presumably much safer and more effective.

The mid-1960s also saw the publication of a series of drug utilization studies. 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 Chapter 27).

With all of these developments, the 1960s can be thought to have marked the beginning of the field of pharmacoepidemiology.

Despite the more stringent process for drug regulation, the late 1960s, 1970s, 1980s, and especially the 1990s and 2000s have seen a series of major adverse drug reactions. Subacute myelo-optic-neuropathy (SMON) was found to be caused by clioquinol, a drug marketed in the early 1930s but not discovered to cause this severe neurological reaction until 1970. 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. The mid-1970s saw the discovery of the oculomucocutaneous syndrome caused by practolol, five years after drug marketing. 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 (Chapter 27) by collecting lifetime drug exposure histories from hospitalized patients and using these to perform hospital-based case-control studies (see Chapter 9). 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. The Computerized Online Medicaid Analysis and Surveillance System was first developed in 1977, using Medicaid billing data to perform

pharmacoepidemiology studies (see Chapters 11 and 12). 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 (see Chapter 10). Each of these represented major contributions to the field of pharmacoepidemiology. These and newer approaches are reviewed in Section II of this book.

In 1980, the drug ticrynafen was noted to cause deaths from liver disease. In 1982, benoxaprofen was noted to do the same. Subsequently, the use of zomepirac, another nonsteroidal anti-inflammatory drug, was noted to be associated with an increased risk of anaphylactoid reactions. Serious blood dyscrasias were linked to phenylbutazone. Small intestinal perforations were noted to be caused by a particular slow release formulation of indomethacin. 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 (see Chapter 27). Acute flank pain and reversible acute renal failure were noted to be caused by suprofen. Isotretinoin was almost removed from the US market because of the birth defects it causes. The eosinophilia-myalgia syndrome was linked to a particular brand of L-tryptophan. Triazolam, thought by the Netherlands in 1979 to be subject to a disproportionate number of central nervous system side effects, was discovered by the rest of the world to be problematic in the early 1990s. 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 was restricted from use except for breast reconstruction after mastectomy. 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. 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. An epidemic of deaths from asthma in New Zealand was traced to fenoterol, and later data suggested that similar, although smaller, risks might be present with other beta-agonist inhalers. The possibility was raised of cancer from depot-medroxyprogesterone, resulting in initial refusal to allow its marketing for contraception in the US, multiple studies, and ultimate approval. Arrhythmias were linked to the use of the antihistamines terfenadine and astemizole. Hypertension, seizures, and strokes were noted from postpartum use of bromocriptine. Multiple different adverse reactions were linked to temafloxacin. Other examples include liver

toxicity from amoxicillin-clavulanic acid; liver toxicity from bromfenac; cancer, myocardial infarction, and gastrointestinal bleeding from calcium channel blockers; arrhythmias with cisapride interactions; 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; anaphylactic reactions from vitamin K; liver toxicity from troglitazone; and intussusception from rotavirus vaccine.

More recently, drug crises have occurred due to allegations of ischemic colitis from alosetron; rhabdomyolysis from cerivastatin; bronchospasm from rapacuronium; torsade de pointes from ziprasidone; hemorrhagic stroke from phenylpropanolamine; arthralgia, myalgia, and neurologic conditions from Lyme vaccine; multiple joint and other symptoms from anthrax vaccine; myocarditis and myocardial infarction from smallpox vaccine; and heart attack and stroke from rofecoxib. Twenty-two different prescription drug products have been removed from the US market since 1980 alone—alosetron (2000), astemizole (1999), benoxaprofen (1982), bromfenac (1998), cerivastatin (2001), cisapride (2000), dexfenfluramine (1997), encainide (1991), fenfluramine (1998), flosequin (1993), grepafloxin (1999), mibefradil (1998), nomifensine (1986), phenylpropanolamine (2000), rapacuronium (2001), rofecoxib (2004), suprofen (1987), terfenadine (1998), temafloxacin (1992), ticrynafen (1980), troglitazone (2000), and zomepirac (1983) (see Chapter 6).

The licensed vaccines against rotavirus and Lyme were also recently withdrawn because of safety concerns (see Chapter 27). Between 1990 and 2004, at least 13 non-cardiac drugs were subject to significant regulatory actions because of cardiac concerns, including astemizole, cisapride, droperidol, grepafloxacin, halofantrine, pimozide, rofecoxib, sertindole, terfenadine, terodiline, thioridazine, vevacetylmethadol, and ziprasidone.

In some of these examples, the drug was never convincingly linked to the adverse reaction. However, many of these discoveries 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.

Each of these was a serious but uncommon drug effect, and these and other serious but uncommon drug effects 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.

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 Section III of this book.

Recent years have seen increasing use of these data resources and new methodologies, with continued and even growing concern about 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, and the International Society for Pharmacoepidemiology issued, in 1996, Guidelines for Good Epidemiology Practices for Drug, Device, and Vaccine Research in the United States, which was very recently updated. In the late 1990s, pharmacoepidemiologic research has been increasingly hampered by concerns about patient confidentiality (see also Chapter 19).

Organizationally, in the US, the Prescription Drug User Fee Act (PDUFA) of 1992 allowed the US FDA to charge manufacturers a fee for reviewing New Drug Applications. This provided additional resources to the FDA, and greatly accelerated the drug approval process. New rules in the US, and in multiple other countries, now permit direct-to-consumer advertising of prescription drugs. The result is a system where more than 330 new medications were approved by the FDA in the 1990s. Each drug costs \$300–500 million to develop; drug development cost the pharmaceutical industry a total of \$24 billion in 1999 and \$32 billion in 2002.

Yet, funds from the PDUFA of 1992 were initially prohibited from being used for drug safety regulation. In 1998, whereas 1400 FDA employees worked with the drug approval process, only 52 monitored safety; the FDA spent only \$2.4 million in extramural safety research. This has coincided with the growing numbers of drug crises cited above. With the passage of PDUFA III, however, this is markedly changing (see Chapter 6). As another measure of drug safety problems, the FDA's new MedWatch program of collecting spontaneous reports of adverse reactions (see Chapter 7) now issues monthly notifications of label changes, and as of mid-1999, 20–25 safety-related label changes are being made every month. According to a study by the US Government Accounting Office,

51% of approved drugs have serious adverse effects not detected before approval. Further, there is recognition that the initial dose recommended for a newly marketed drug is often incorrect, and needs monitoring and modification after marketing.

Recently, 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, there has been increased concern about reliance solely on nonexperimental methods to study drug safety after marketing, and we are beginning to see the use of massive randomized clinical trials as part of postmarketing surveillance (see Chapter 20).

There is also increasing recognition that most of the risk from most drugs to most patients occurs from known reactions to old drugs. Yet, nearly all of the efforts by the FDA and other regulatory bodies are devoted to discovering rare unknown risks from new drugs. In response, there is growing concern, in Congress and among the US public at least, that perhaps the FDA is now approving drugs too *fast*. There are also calls for the development of an independent drug safety board, analogous to the National Transportation Safety Board, with a mission much wider than the FDA's regulatory mission, to complement the latter. For example, such a board could investigate drug safety crises such as those 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.

As an attempt to address the kinds of questions which until now have not been addressed, the US Agency for Healthcare Research and Quality (AHRQ) has funded seven Centers for Education and Research on Therapeutics (CERTs). Discussed more in Chapter 6, the CERTs program seeks to improve health care and patient safety. It has identified specific roles that include: (a) development and nurturing of public-private partnerships to facilitate research on therapeutics; (b) support and encouragement of research on therapeutics likely to get translated into policy or clinical practice; (c) development of educational modules and dissemination strategies to increase awareness of the benefits and risks of pharmaceuticals; and (d) creation of a national information resource on the safe and effective use of therapeutics. Activities include the conduct of research on therapeutics, specifically exploring new uses of drugs, ways to improve the effective uses of drugs, and risks associated with new uses or combinations of drugs. The CERTs also develop educational modules and materials for disseminating the findings from their research, consistent with

their overarching mission to become a national resource for people seeking information about medical products. The CERTs strive to seek public and private sector cooperation to facilitate these efforts.

Another new 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: (a) "even apparently single events or errors are due most often to the convergence of multiple contributing factors," (b) "preventing errors and improving safety for patients requires a systems approach in order to modify the conditions that contribute to errors," and (c) "the problem is not bad people; the problem is that the system needs to be made safer." 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 adverse drug events (ADEs), prolonging hospital stays by 2 days, and increasing costs by \$2000–2600 per patient. Over 7000 US deaths were attributed to medication errors in 1993. Although these estimates have been disputed, the overall importance of reducing these errors has not been questioned. In recognition of this problem, the AHRQ has launched a major new grant program of over 100 projects, 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 27. A recent CERT paper called for a systematic review of the entire drug risk assessment process, perhaps as a study by the US Institute of Medicine. That study is underway, at least in part in response to the circumstances surrounding the withdrawal of rofecoxib.

Finally, 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, there are many initiatives underway, ranging from new FDA requirements for risk management plans, to a new FDA Drug Safety and

Risk Management Advisory Committee. More information is provided in Chapters 6 and 27.

THE CURRENT DRUG APPROVAL PROCESS

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 and 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 2). 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 (see Chapter 2 for a discussion of confidence intervals). At the other extreme, a total of 500 patients would allow one to be 95% certain of detecting any adverse reactions which occur in 6 or more patients out of every 1000 exposed. Adverse reactions which 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 3.

POTENTIAL CONTRIBUTIONS OF PHARMACOEPIDEMIOLOGY

The potential contributions of pharmacoepidemiology are only beginning to be realized, as the field is relatively new. However, some contributions are already apparent (see Table 1.1). In fact, since the early 1970s the FDA has required postmarketing research at the time of approval for about one third of drugs. In this section we will first review the potential for pharmacoepidemiology studies to supplement the information available prior to marketing, and then review the new types of information obtainable from postmarketing pharmacoepidemiology studies but not obtainable prior to drug marketing. Finally, we will review the general, and probably most important, potential contributions such studies can make. In each case, the relevant information available from premarketing studies will be briefly examined first, to clarify how postmarketing studies can supplement this information.

SUPPLEMENTARY INFORMATION

Premarketing studies of drug effects are necessarily limited in size. After marketing, nonexperimental epidemiologic studies can be performed, evaluating the effects of drugs administered as part of ongoing medical care. These allow the cost-effective accumulation of much larger numbers of

Table 1.1. Potential contributions of pharmacoepidemiology

-
(A)	Information which supplements the information available from premarketing studies—better quantitation of the incidence of known adverse and beneficial effects
	(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
-

patients than those studied prior to marketing, resulting in a more precise measurement of the incidence of adverse and beneficial drug effects (see Chapter 3). For example, at the time of drug marketing, prazosin was known to cause a dose-dependent first dose syncope, but the FDA requested the manufacturer to conduct a postmarketing surveillance study in the US to quantitate its incidence more precisely. 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. As noted above, the increased sample size available after marketing also permits a more precise determination of the correct dose to be used.

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.

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 examine the effects of differences in drug regimen, compliance, etc. 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.

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.

NEW TYPES OF INFORMATION NOT AVAILABLE FROM PREMARKETING STUDIES

As mentioned above, premarketing studies are necessarily limited in size. 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.

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.

The patterns of physician prescribing and patient drug utilization often cannot be predicted prior to marketing, despite pharmaceutical manufacturers' best attempts to predict in 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 Chapter 27).

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 pharmacoepidemiology studies.

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 have been applied to evaluate the cost implications of drug use. 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. 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 Chapter 22).

GENERAL CONTRIBUTIONS OF PHARMACOEPIDEMOLOGY

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 pharmacoepidemiology 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, so, whether it was negligent in its behavior.

Key Points

- *Pharmacoepidemiology* is the study of the use of and the effects of drugs in large numbers of people. It uses the methods of epidemiology to study the content area of clinical pharmacology.
- The history of pharmacoepidemiology is a history of increasingly frequent accusations about adverse drug reactions, often arising out of the spontaneous reporting system, followed by formal studies proving or disproving those associations.
- The drug approval process is inherently limited, so it cannot detect, before marketing, adverse effects that are uncommon, delayed, unique to high risk populations, due to misuse of the drugs by prescribers or patients, etc.
- Pharmacoepidemiology can contribute information about drug safety and effectiveness that is not available from premarketing studies.

SUGGESTED FURTHER READINGS

- Califf RM. The need for a national infrastructure to improve the rational use of therapeutics. *Pharmacoepidemiol Drug Saf* 2002; **11**: 319–27.
- Caranasos GJ, Stewart RB, Cluff LE. Drug-induced illness leading to hospitalization. *JAMA* 1974; **228**: 713–17.
- Cluff LE, Thornton GF, Seidl LG. Studies on the epidemiology of adverse drug reactions. I. Methods of surveillance. *JAMA* 1964; **188**: 976–83.
- 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.
- Erslev AJ, Wintrobe MM. Detection and prevention of drug induced blood dyscrasias. *JAMA* 1962; **181**: 114–19.
- Geiling EMK, Cannon PR. Pathogenic effects of elixir of sulfanilimide (diethylene glycol) poisoning. *JAMA* 1938; **111**: 919–26.
- Guidelines For Good Pharmacoepidemiology Practices. *Pharmacoepidemiol Drug Saf* 2005; **14**: 589–95.
- 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.
- Joint Commission on Prescription Drug Use. *Final Report*. Washington, DC, 1980.
- 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.
- Kono R. Trends and lessons of SMON research. In: Soda T, ed., *Drug-Induced Sufferings*. Princeton, NJ: Excerpta Medica, 1980; p. 11.
- 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.
- Lenz W. Malformations caused by drugs in pregnancy. *Am J Dis Child* 1966; **112**: 99–106.
- Meyler L. *Side Effects of Drugs*. Amsterdam: Elsevier, 1952.
- Miller RR, Greenblatt DJ. *Drug Effects in Hospitalized Patients*. New York: John Wiley & Sons, 1976.
- Rawlins MD, Thompson JW. Pathogenesis of adverse drug reactions. In: Davies DM, ed., *Textbook of Adverse Drug Reactions*. Oxford: Oxford University Press, 1977; p. 44.
- 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.
- 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.
- Wallerstein RO, Condit PK, Kasper CK, Brown JW, Morrison FR. Statewide study of chloramphenicol therapy and fatal aplastic anemia. *JAMA* 1969; **208**: 2045–50.
- Wright P. Untoward effects associated with practolol administration. Oculomucocutaneous syndrome. *BMJ* 1975; **1**: 595–8.

2

Study Designs Available for Pharmacoepidemiology Studies

Edited by:

BRIAN L. STROM

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

Pharmacoepidemiology applies the methods of epidemiology to the content area of clinical pharmacology. 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. 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 how likely 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 pharmacoepidemiology 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. Finally,

Chapter 4 will review basic principles of clinical pharmacology, the content area of pharmacoepidemiology, in a similar manner.

OVERVIEW OF THE SCIENTIFIC METHOD

The scientific method is a three-stage process (see Figure 2.1). In the first stage, one selects a group of subjects for study. 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

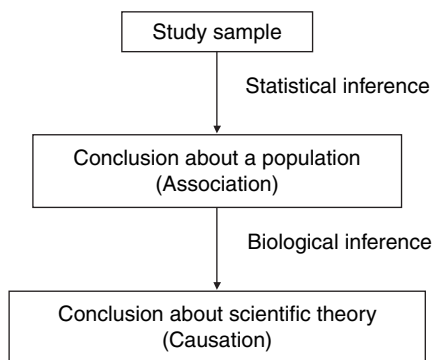


Figure 2.1. Overview of the scientific method.

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

At this point, one would be tempted to make a generalization that enalapril lowers blood pressure in middle-aged hypertensive men. However, one must explore whether this observation could have occurred simply by chance, i.e., 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,” 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 2.1 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 obvious 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 resulting from such errors.

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 2.1). 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*, i.e., a spurious or false association. This can occur by either of two mechanisms: chance or bias. Chance is unsystematic, or random, variation. The purpose of statistical testing in

Table 2.1. Types of association 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) |

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. 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 a 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.

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 16 for more discussion about biases in pharmacoepidemiology studies.)

Third, one can have an indirect, or confounded, association. A *confounding variable*, or *confounder*, is a variable other than the risk factor and outcome under study which is related independently to both the risk factor and the outcome variable 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,

Table 2.2. Approaches to controlling confounding

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

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 2.2. (See Chapters 16 and 21 for more discussion about confounding in pharmacoepidemiology 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 in 1965, but have been described in various forms since, each with some modification. Probably the best known description of them was in the US Public Health Service's first Surgeon General's Report on Smoking and Health, published in 1964. These criteria are presented in Table 2.3, 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.

The first criterion listed in Table 2.3 is *coherence with existing information* or *biological plausibility*. This refers

Table 2.3. 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
-

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. 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 2.3 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. 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 2.3 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 stressed when it attempted to make the claim that cigarette smoking had 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 2.3 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 methodology 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.

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

Table 2.4. Advantages and disadvantages of epidemiologic study designs

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

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.

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 2.4 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, 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 associa-

tions 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 not be ethical or logistically feasible to randomly allocate individuals to smoke or not to smoke and expect them to follow that for 20 years to observe the outcome in each group.

The issue of causation is discussed more in Chapters 7 and 8 as it relates to the process of spontaneous reporting of adverse drug reactions, and in Chapter 17 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 2.4 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. 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–rechallenge situation is relatively uncommon, however, as physicians generally will avoid exposing a patient to a drug if the patient experienced an adverse reaction to it in the past. This issue is discussed in more detail in Chapters 7, 8, and 17.

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 when observed 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. The “Phase IV” postmarketing surveillance study of cimetidine 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. 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, sometimes 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.

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. This is certainly consistent with the increasing rates of cigarette smoking observed in women until the mid-1960s, 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.

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 3). For example, the classic study of diethylstilbestrol and clear cell vaginal adenocarcinoma required only 8 cases and 40 controls, 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, i.e., by recreating events that happened in the past. Information on past exposure to potential risk factors is generally obtained by abstracting medical records or by administering questionnaires or interviews. As such, case–control studies are subject to limitations in the validity of retrospectively collected exposure information. In addition, the proper selection of controls can be a challenging task, and inappropriate 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 pharmacoepidemiology 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. 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 2.2). Patients are recruited into case-control studies based on the presence or absence of a disease, and their antecedent exposures are then studied. Patients are recruited into cohort studies based on the presence or absence of an exposure, and their subsequent disease course is then studied.

Cohort studies have the major advantage of being free of the big 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, particularly a prospective one, 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 3). 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 2.2, both case-control and cohort studies are intended to provide the same basic information; the difference is how this information is collected. The key measure of effect that the exposure has on the outcome 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 usually 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

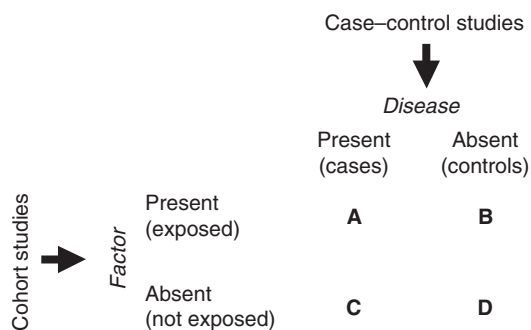


Figure 2.2. Cohort and case-control studies provide similar information, but approach data collection from opposite directions. (Reprinted with permission from Strom BL. Medical databases in post-marketing drug surveillance. *Trends in Pharmacological Sciences* 1986; 7: 377–80.)

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 ten-fold increase in risk that is also statistically significant. A study that yields a relative risk (95% confidence interval) of 10.0 (1.1–100) says little, other than 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. However, the risk of myocardial infarction in non-smoking women in their 20s is so low, that even a five-fold 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. Confidence intervals can be calculated around excess risks as well, and would be interpreted analogously.

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.

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 pharmacoepidemiology studies mainly for supplementary studies of drug efficacy. 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, there has been increased concern about reliance solely on nonexperimental methods to study drug safety after marketing, and we are beginning to see the use of massive randomized clinical trials as part of postmarketing surveillance (see Chapter 20).

DISCUSSION

Thus, a series of different study designs are available (Table 2.4), 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 *nonexperimental study designs*, in order to differentiate them from experimental studies. In nonexperimental study designs the investigator does not control the therapy, but simply observes and evaluates the results of ongoing medical care. Case reports, case series, and analyses of secular trends have

Table 2.5. 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

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 2.5). From the perspective of how subjects are recruited into the study, case-control studies can be contrasted with cohort studies. Specifically, case-control studies select subjects into the study based on the presence or absence of a disease, while cohort studies select subjects into the study based on the presence or absence of an exposure. From this perspective, randomized clinical trials can be viewed as a subset of cohort studies, a type of cohort study in which the investigator controls the allocation of treatment, rather than simply observing ongoing medical care. From the perspective of timing, data can be collected *prospectively*, that is simultaneously with the events under study, or *retrospectively*, that is after the events under study had already developed. In the latter situation, one recreates events that happened in the past using medical records, questionnaires, or interviews. Data can also be collected using *cross-sectional studies*, studies that have no time sense, as they examine only one point in time. In principle, either cohort or case-control studies can be performed using any of these time frames, although prospective case-control studies are unusual. Randomized clinical trials must be prospective, as this is the only way an investigator can control the therapy received.

The terms presented in this chapter, which are those that will be used throughout the book, are probably the terms used by a majority of epidemiologists. Unfortunately, however, other terms have been used for most of these study designs as well. Table 2.5 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 2.5, however, in the past some authors used the term “retrospective study” to refer to a case-control study and 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 nonexperimental 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 2.4 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. 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. 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 and analyses of secular trends. 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.

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 (1965):

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.

Key Points

- There are four basic types of association that one can have: no association, artifactual association (from chance or bias), an indirect association (from confounding), or a true association.
- There are a series of criteria for the causal nature of an association, which assist in making the subjective judgment about whether a given association is likely to be causal. Included are biological plausibility, consistency, time sequence, specificity, and quantitative strength.
- Study design options, in hierarchical order of progressively harder to perform but more convincing, are: case reports, case series, analyses of secular trends, case-control studies, retrospective cohort studies, prospective cohort studies, and randomized clinical trials.

SUGGESTED FURTHER READINGS

- Ahlbom A, Norell S. *Introduction to Modern Epidemiology*, 2nd edn. Chestnut Hill, MA: Epidemiology Resources, 1990.
- Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials*, 3rd edn. Baltimore, MD: Williams and Wilkins, 1996.
- Friedman G. *Primer of Epidemiology*, 3rd edn. New York: McGraw-Hill, 1994.
- Gordis L. *Epidemiology*, 2nd edn. Philadelphia, PA: Saunders, 2000.
- Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring JR. *Medical Epidemiology*, 3rd edn. New York: McGraw-Hill, 2001.
- Hennekens CH, Buring JE. *Epidemiology in Medicine*, Boston, MA: Little, Brown, 1987.
- Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; **58**: 295–300.
- Hulley SB, Cummings SR. *Designing Clinical Research: An Epidemiologic Approach*, Baltimore, MD: Williams and Wilkins, 1988.
- Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. *Designing Clinical Research: An Epidemiologic Approach*, 2nd edn. Baltimore, MD: Williams and Wilkins, 2001.
- Kelsey JL, Thompson WD, Evans AS. *Methods in Observational Epidemiology*, New York: Oxford University Press, 1986.
- Lilienfeld DE, Stolley P. *Foundations of Epidemiology*, 3rd edn. New York: Oxford University Press, 1994.
- MacMahon B, Pugh TF. *Epidemiology: Principles and Methods*, Boston, MA: Little, Brown, 1970.
- Mausner JS, Kramer S. *Epidemiology: An Introductory Text*, 2nd edn. Philadelphia, PA: Saunders, 1985.
- Rothman KJ. *Epidemiology: An Introduction*, New York: Oxford University Press, 2002.
- Rothman KJ, Greenland S. *Modern Epidemiology*, 2nd edn. Philadelphia, PA: Lippincott-Raven, 1998.
- Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*, 2nd edn. Boston, MA: Little, Brown, 1991.
- Schuman SH. *Practice-Based Epidemiology*, New York: Gordon and Breach, 1986.
- 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.
- Weiss N. *Clinical Epidemiology: The Study of the Outcome of Illness*, 2nd edn. New York: Oxford University Press, 1996.

3

Sample Size Considerations for Pharmacoepidemiology Studies

Edited by:

BRIAN L. STROM

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 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 pharmacoepidemiology 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 Section II 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. 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.

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 pharmacoepidemiology 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 3.1).

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, i.e., 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, i.e., 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.

The fourth variable one needs to specify is the expected incidence of the outcome 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.

Table 3.1. 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 odds ratio 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 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 ratio of the number of controls 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. This is done for reasons of validity, not 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 each comparison 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 (1974):

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

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 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 Appendix A, which were calculated using this formula. In Tables A1–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 A.5–A.8 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.

For example, what if one wanted to investigate a new nonsteroidal anti-inflammatory drug that is about to be marketed, but premarketing data raised questions about possible hepatotoxicity? This would presumably be studied using a cohort study design and, depending upon the values chosen for α , β , the incidence of the disease in the unexposed, 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 3.2). 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 A.1 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 3.2 and Table A.5). If one wanted to minimize the number of exposed subjects needed for the study, one could include

Table 3.2. 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	31 483	31 483
	0.001	0.05 (2-tailed)	0.2	2	1	23 518	23 518
	0.001	0.05 (2-tailed)	0.2	2	4	13 402	53 608
	0.001	0.05 (1-tailed)	0.2	2	4	10 728	42 912
	0.001	0.05 (2-tailed)	0.1	4	1	5823	5823
	0.001	0.05 (2-tailed)	0.2	4	1	4350	4350
	0.001	0.05 (2-tailed)	0.2	4	4	2253	9012
	0.001	0.05 (1-tailed)	0.2	4	4	1829	7316
Cholestatic jaundice	0.0001	0.05 (2-tailed)	0.1	2	1	315 268	315 268
	0.0001	0.05 (2-tailed)	0.2	2	1	235 500	235 500
	0.0001	0.05 (2-tailed)	0.2	2	4	134 194	536 776
	0.0001	0.05 (1-tailed)	0.2	2	4	107 418	429 672
	0.0001	0.05 (2-tailed)	0.1	4	1	58 376	58 376
	0.0001	0.05 (2-tailed)	0.2	4	1	43 606	43 606
	0.0001	0.05 (2-tailed)	0.2	4	4	22 572	90 288
	0.0001	0.05 (1-tailed)	0.2	4	4	18 331	73 324

up to four controls for each exposed subject (Table 3.2 and Table A.8). This would result in a sample size of 13 402, with four times as many controls, a total of 67 010 subjects. Finally, if one considers it inconceivable that this new drug could *protect* against liver disease and one is not interested in that outcome, then one might use a one-tailed α , resulting in a somewhat lower sample size of 10 728, again with four times as many controls. Much smaller sample sizes are needed to detect relative risks of 4.0 or greater; these are also presented in Table 3.2.

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 315 268 subjects in each study group.

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

SAMPLE SIZE CALCULATIONS FOR CASE-CONTROL STUDIES

The approach to calculating sample sizes for case-control studies is similar to the approach for cohort studies. Again, there are five variables that need to be specified (see Table 3.1). 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 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, 1974):

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

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,

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

and

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

Again, a series of tables that provide sample sizes for case-control studies is presented in Appendix A. In Tables A.9–A.12, we have assumed an α (two-tailed) of 0.05, a beta of 0.1 (90% power), and control to case ratios of 1 : 1, 2 : 1, 3 : 1, and 4 : 1, respectively. Tables A.13–A.16 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 nonsteroidal anti-inflammatory drug that is about to

Table 3.3. 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	31 588	31 588
	0.001	0.05 (2-tailed)	0.2	2	1	23 596	23 596
	0.001	0.05 (2-tailed)	0.2	2	4	13 449	53 796
	0.001	0.05 (1-tailed)	0.2	2	4	10 765	43 060
	0.001	0.05 (2-tailed)	0.1	4	1	5856	5856
	0.001	0.05 (2-tailed)	0.2	4	1	4375	4375
	0.001	0.05 (2-tailed)	0.2	4	4	2266	9064
	0.001	0.05 (1-tailed)	0.2	4	4	1840	7360
Phenylbutazone	0.0001	0.05 (2-tailed)	0.1	2	1	315 373	315 373
	0.0001	0.05 (2-tailed)	0.2	2	1	235 579	235 579
	0.0001	0.05 (2-tailed)	0.2	2	4	134 240	536 960
	0.0001	0.05 (1-tailed)	0.2	2	4	107 455	429 820
	0.0001	0.05 (2-tailed)	0.1	4	1	58 409	58 409
	0.0001	0.05 (2-tailed)	0.2	4	1	43 631	43 631
	0.0001	0.05 (2-tailed)	0.2	4	4	22 585	90 340
	0.0001	0.05 (1-tailed)	0.2	4	4	18 342	73 368

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 3.3). 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 at 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 3.3 and Table A.13). If one wanted to minimize the number of diseased subjects needed for the study, one could include up to four controls for each exposed subject (Table 3.3 and Table A.16). 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 3.3.

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 patients, 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 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.

SAMPLE SIZE CALCULATIONS FOR CASE SERIES

As described in Chapter 2, the utility of case series in pharmacoepidemiology is limited, as the absence of a

control group makes causal inference difficult. Despite this, however, this is a design that has been used repeatedly. There are scientific questions that can be addressed using this design, and the collection of a control group equivalent in size to the case series would add considerable cost to the study. Case series are usually used in pharmacoepidemiology to quantitate better the incidence of a particular disease in patients exposed to a newly marketed drug. For example, in the "Phase IV" postmarketing drug surveillance study conducted for prazosin, the investigators collected a case series of 10 000 newly exposed subjects recruited through the manufacturer's sales force, to quantitate better the incidence of first-dose syncope, which was a well-recognized adverse effect of this drug. 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.

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 A.17 in Appendix A presents a table useful for making this calculation. In order to apply this table, one first calculates the incidence rate observed from the study's results, that is the number of subjects who develop the disease of interest during the specified time interval, divided by the total number of individuals in the population at risk. For example, if three cases of liver disease were observed in a population of 1000 patients exposed to a new nonsteroidal anti-inflammatory drug during a specified period of time, the incidence would be 0.003. The number of subjects who develop the disease is the "Observed number on which estimate is based (n)" in Table A.17. In this example, it is 3. The lower boundary of the 95% confidence interval for the incidence rate is then the corresponding "Lower limit factor (L)" multiplied by the observed incidence rate. In 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. Specifically, if no events of a particular type 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 3 in 3000 subjects, or the events have 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 3 in 10 000 exposed individuals, or that they have an incidence rate of less than 0.0003. In other words, events not detected in the study may occur less often than 1 in 3333 subjects.

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. It is usually better to have a somewhat larger than expected sample size than the minimum, anyway, 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.

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, i.e., 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, i.e., 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.

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 pharmacoepidemiology 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 pharmacoepidemiology 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 Section II of this book.

Second, because of the development of these new techniques, pharmacoepidemiology 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 2). Subtle findings should not be ignored, but should be interpreted with caution.

Key Points

- Premarketing studies of drugs are inherently limited in size, meaning larger studies are needed after marketing in order to detect less common drug effects.
- For a cohort study, the needed sample size is determined by specifying the type I error one is willing to tolerate, the type II error one is willing to tolerate, the smallest relative risk that one wants to be able to detect, the expected incidence of the outcome of interest in the unexposed control group, and the ratio of the number of unexposed control subjects to be included in the study to the number of exposed study subjects.
- For a case-control study, the needed sample size is determined by specifying the type I error one is willing to tolerate, the type II error one is willing to tolerate, the smallest odds ratio that one wants to be able to detect, the expected prevalence of the exposure of interest in the undiseased control group, and the ratio of the number of undiseased control subjects to be included in the study to the number of diseased study subjects.
- As a rule of thumb, if no events of a particular type are observed in a study of X individuals, then one can be 95% certain that the event occurs no more often than $3/X$.

SUGGESTED FURTHER READINGS

- Cohen J. *Statistical Power Analysis for the Social Sciences*. New York: Academic Press, 1977.
- Gifford LM, Aeugle ME, Myerson RM, Tannenbaum PJ. Cimetidine postmarket outpatient surveillance program. *JAMA* 1980; **243**: 1532–5.
- Graham RM, Thornell IR, Gain JM, Bagnoli C, Oates HF, Stokes GS. Prazosin: the first dose phenomenon. *BMJ* 1976; **2**: 1293–4.
- 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.
- Joint Commission on Prescription Drug Use. *Final Report*. Washington, DC, 1980.
- Makuch RW, Johnson MF. Some issues in the design and interpretation of “negative” clinical trials. *Arch Intern Med* 1986; **146**: 986–9.
- Schlesselman JJ. Sample size requirements in cohort and case-control studies of disease. *Am J Epidemiol* 1974; **99**: 381–4.
- Stolley PD, Strom BL. Sample size calculations for clinical pharmacology studies. *Clin Pharmacol Ther* 1986; **39**: 489–90.
- Young MJ, Bresnitz EA, Strom BL. Sample size nomograms for interpreting negative clinical studies. *Ann Intern Med* 1983; **99**: 248–51.

4

Basic Principles of Clinical Pharmacology Relevant to Pharmacoepidemiology Studies

Edited by:

SEAN HENNESSY

Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia,
Pennsylvania, USA.

INTRODUCTION

Clinical pharmacology comprises all aspects of the scientific study of medicinal drugs in humans. Its overall objective is to provide the knowledge base needed to ensure rational drug therapy. In addition to studying biologic effects of drugs, clinical pharmacology includes the study of non-pharmacologic (e.g., economic and social) determinants and effects of medication use. The development of clinical pharmacology had its roots in the so-called “drug explosion” that occurred between the 1930s and 1960s, which was marked by a pronounced escalation of the rate at which new drugs entered the markets of economically developed nations. With this rapid expansion of the therapeutic armamentarium came the need for much more information regarding the effects and optimal use of these agents, which spurred the growth of clinical pharmacology as a scientific discipline.

Some would define an additional related discipline, *pharmacotherapeutics*, which 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.

Clinical pharmacology tries to explain the response to drugs in *individuals*, while pharmacoepidemiology is concerned with measuring and explaining variability in outcome of drug treatment in *populations*. However, there is great overlap in the scope of the two disciplines and many clinical pharmacologists are heavily involved in pharmacoepidemiologic research. Pharmacoepidemiology is the application of epidemiologic methods to the subject matter of clinical pharmacology. Of course, neither approach would be justified if responses to drugs were totally predictable. From this perspective, the origins of pharmacoepidemiology can be seen clearly in the disciplines of clinical pharmacology and pharmacotherapeutics.

In epidemiologic studies of non-drug exposures, it is frequently assumed that the amount and duration of exposure is proportional to the risk of the outcome. For instance, the risk of a stroke or heart attack is often presumed to

increase in proportion both to the level of a risk factor, such as elevated blood pressure or blood cholesterol, and to the length of time the risk factor has been present. Likewise, duration of exposure to carcinogens (e.g., cigarette smoke) is sometimes assumed to be linearly related to the level of risk. On occasion, these proportionality assumptions hold true in pharmacoepidemiology. For instance, the risk of endometrial cancer increases in direct proportion to the duration of exposure to estrogens. In other situations, proportionality assumptions are invalid, as is the case with rashes, hepatic reactions, and hematologic reactions to drugs, which often occur in the first few weeks of treatment, the risk declining thereafter. These apparently declining risks may be an artifact of the epidemiologic phenomenon known as “depletion of susceptibles” (where long-term users of a drug class tend to be those who are tolerant of the drug’s effects), and/or they may be due to a number of biologic factors that are unique to the ways in which drugs elicit responses, are handled by the body, and are used in clinical practice.

Exposure to a drug is never a completely random event, as individuals who receive a drug almost always differ from those not receiving it. The circumstances leading to a patient receiving a particular drug in a particular dose, at a particular time, are complex and relate to the patient’s health care behavior and use of services, the severity and nature of the condition being treated, and the perceived advantages of a drug in a specific setting. For many conditions, physicians alter or titrate the dose of a drug against a response, and will tend to switch medications in the case of non-response. Consequently, the choice of a drug and dose may be determined by factors that are themselves related to the outcome under study. In other words, the association between the drug and the outcome of interest may be confounded by the indication for the drug or other related features (see also Chapter 21). Because of the high probability of confounding beyond that which can be controlled for using measured variables (i.e., residual confounding), pharmacoepidemiologists tend to be cautious about the interpretation of weak associations between drug exposure and outcomes.

When interpreting pharmacoepidemiology studies, it is important to realize that relationships exist between drug response and various biologic and sociologic factors, and to attempt to explore the reasons for them. The discipline of clinical pharmacology has provided us with explanations for some of these variations in response to important drugs, and knowledge of these is necessary when conducting or interpreting pharmacoepidemiology studies.

This chapter is intended to introduce readers to some of the core concepts of clinical pharmacology. Obviously,

a single book chapter cannot convey the entire discipline; many general and topic-specific clinical pharmacology textbooks exist which accomplish this. The emphasis of this chapter will be on concepts that are likely to be important in conducting and understanding pharmacoepidemiologic research. In particular, one of the most important areas of study within clinical pharmacology that is inherently amenable to the use of epidemiologic methods is the variability of drug response that exists across the population. The following sections present some of the central concepts of clinical pharmacology that are important to the pharmacoepidemiologist who is attempting to understand differences in the population with regard to the effects of drugs. Specifically, this chapter will discuss the nature of drugs, the mechanisms of drug action, the concept of drug potency, the role of pharmacodynamics and pharmacokinetics (including genetic factors that influence these functions), and the importance of human behavior in explaining variability in drug effects.

THE NATURE OF DRUGS

A drug may be defined as any exogenously administered substance that exerts a physiologic effect. Taken as a group, drugs vary greatly with regard to their molecular structure. For example, interferon α -2a is an intricate glycoprotein, while potassium chloride is a simple salt containing only two elements. Most drugs are intermediate in complexity, and produce their pharmacologic response by exerting a chemical or molecular influence on one or more cell constituents.

Typically, the active drug component of a tablet, capsule, or other pharmaceutical dosage form accounts for only a small percentage of the total mass and volume. The remainder is composed of excipients (such as binders, diluents, lubricants, preservatives, coloring agents, and sometimes flavoring) that are chosen, among other concerns, because they are believed to be pharmacologically inert. This is relevant to the pharmacoepidemiologist because a drug product’s ostensibly inactive ingredients can sometimes produce effects of their own. For example, benzyl alcohol, which is commonly used as a preservative in injectable solutions, has been implicated as the cause of a toxic syndrome that has resulted in the deaths of a number of infants.

Also of potential concern to the pharmacoepidemiologist is the fact that, over time, a pharmaceutical product can be reformulated to contain different excipients. Furthermore, because of the marketing value of established proprietary drug product names, non-prescription products are

sometimes reformulated to contain different active ingredients, and then continue to be marketed under their original brand name. This is potentially of concern to any pharmacoepidemiologist interested in studying the effects of non-prescription drugs. It also is a potential source of medication errors (see also Chapter 27).

MECHANISMS OF DRUG ACTION

Pharmacology seeks to characterize the actions of drugs at many different levels of study, such as the organism, organ, tissue, cell, cell component, and molecular levels. On the macromolecular level, most drugs elicit a response through interactions with specialized proteins such as enzymes and cell surface receptors. While drug molecules may be present within body fluids either in their free, native state, or bound to proteins or other constituents, it is typically the free or unbound fraction that is available to interact with the target proteins, and is thus important in eliciting a response.

Enzymes are protein catalysts, or molecules that permit certain biochemical reactions to occur more rapidly. By directly inhibiting an enzyme, a drug may block the formation of its product. For instance, inhibition of angiotensin-converting enzyme blocks the conversion of angiotensin I to its active form, angiotensin II, resulting in a fall in arteriolar resistance that is beneficial to individuals with hypertension or congestive heart failure. Other drugs block ion channels, and consequently alter intracellular function. For example, calcium channel blocking drugs reduce the entry of calcium ions into smooth muscle cells, thereby inhibiting smooth muscle contraction, dilating blood vessels, and so reducing arteriolar resistance.

Alternatively, drugs may interact with specialized receptors on the cell surface, which activate a subsequent intracellular signaling system, ultimately resulting in changes in the intracellular milieu. For instance, drugs that bind to and activate β_2 -adrenoceptors (β_2 -agonists) in the pulmonary airways increase intracellular cyclic adenosine monophosphate concentrations and activate protein kinases, resulting in smooth muscle relaxation and bronchodilation. Many drugs act through interaction with G-protein-coupled receptors on the surface of cells. These are specialized protein receptors that thread through the double lipid layer in cell membranes and broadcast to the inside of the cell that a drug is on the outside. Other drugs, such as the purine and pyrimidine antagonists that are used in cancer chemotherapy, and the nucleoside analogues that are used in the treatment of HIV and other viral infections, exert their effects by blocking cell replication processes.

DRUG POTENCY

In its pharmacologic usage, the term potency refers to the amount of drug that is required to elicit a given response, and is important when one is comparing two or more drugs that have similar effects. For example, 10 mg morphine has approximately the same analgesic activity as 1.3 mg hydromorphone when both drugs are administered by injection. Thus, we say that 10 mg morphine is approximately “equipotent” to 1.3 mg hydromorphone, and that hydromorphone is approximately 7.7 times as potent as morphine ($10/1.3 = 7.7$). As an aside, there is sometimes a tendency to equate potency with “effectiveness,” yielding the misconception that because one drug is more potent than its alternative, it is therefore more effective. This view is fallacious. As the active drug component typically accounts for only a small portion of a pharmaceutical dosage form, the amount of drug that can be conveniently delivered to the patient is rarely at issue; if need be, the dose can simply be increased. Milligram potency is rarely an important consideration in therapeutic drug use, while maximal efficacy (which indicates the maximum effect the drug can exert) is much more important.

On the other hand, drug potency may be important in interpreting pharmacoepidemiology studies. For example, if a particular drug is noted to have a higher rate of adverse effects than other drugs of the same class, it is important to investigate whether this is a result of an intrinsic effect of that drug, or if the drug is being used in clinical practice at a higher dose, relative to its potency, than other drugs of the class. For instance, this may explain some of the apparent differences in risks of serious gastrointestinal complications with individual nonsteroidal anti-inflammatory drugs.

PHARMACODYNAMICS AND PHARMACOKINETICS

Clinical pharmacology can be divided broadly into pharmacodynamics and pharmacokinetics. *Pharmacodynamics* quantifies the response of the target tissues in the body to a given concentration of a drug. *Pharmacokinetics* is the study of the processes of drug absorption, distribution, and elimination from the body. Put simply, pharmacodynamics is concerned with the drug’s action on the body, while pharmacokinetics is concerned with the body’s action on the drug. The combined effects of these processes determine the time course of concentrations of a drug at its target sites and the consequences of the presence of the drug at that concentration. The role of each in contributing to the

variability of drug effects among the population will be discussed in turn.

THE ROLE OF PHARMACODYNAMICS IN DETERMINING VARIABILITY OF DRUG RESPONSE

Compared with most non-drug exposures, there is considerable existing knowledge about the effects of a drug by the time it is marketed. This must be incorporated into the design of new studies that seek to gain further information about that drug's actions. This is true whether the design of the new study is experimental or nonexperimental. Further, there is considerable information about determinants of patients' responses to drugs in general. In this section, we present the effects of genetics, adaptive responses, age, disease states, and concomitant drugs in determining variability in drug response.

GENETIC DETERMINANTS OF HUMAN RESPONSES TO DRUGS

This is the most rapidly developing area of research in clinical pharmacology (also see Chapter 18). The science of genomics helps us understand (and possibly predict) who will respond (or not) to a drug and who will develop serious toxicity at doses that are normally therapeutic in effect. Investigations into genetic determinants have concentrated on three main areas: drug actions, drug transporters, and drug metabolism. The latter two topics are discussed later in the sections on pharmacokinetics. Single nucleoside polymorphisms in genes that code for drug receptors can result in variability in responses to certain drugs. Genetic polymorphisms are differences in the sequence of DNA occurring with a frequency of 1% or more, which can lead to the formation of proteins that do not work properly. For example, polymorphism of the β_2 adrenoceptor leads to lack of response to bronchodilators, and genetic variations in the 5HT_{2a} receptor lead to resistance to the anti-psychotic agent clozapine. Polymorphisms of different genes affecting platelet and endothelial cell function may be associated with an increased risk of thrombosis, and a relative resistance to the anti-thrombotic effects of low dose aspirin. Polymorphisms in response genes may also predict the risk of adverse reactions. Mutations of multiple genes are associated with the "long QT syndrome" and a propensity to ventricular arrhythmia with drugs such as antihistamines, macrolide antibiotics, and cisapride. The affected genes encode cardiac ion channels (K^+ and Na^+), which have a

major role in suppressing arrhythmias initiated by premature beats.

It can be appreciated that gene polymorphisms have potentially important roles in explaining variations in the beneficial and adverse effects to a wide range of drugs. Pharmacoeepidemiology, historically, has estimated the average effects of drugs in populations and the trend has been to pursue efficient designs, particularly through the exploitation of data stored in electronic medical records and administrative databases (see Chapters 11 and 12), which can be linked in order to study the relationship between exposure and outcome. As the science of pharmacogenomics evolves, there is a growing need to incorporate biologic sampling, either through revisiting more efficient *ad hoc* study designs, or through linkage of medical or administrative records to banks of biosamples. These designs will raise significant ethical issues that have not been features of traditional pharmacoeepidemiology studies (see also Chapter 19).

EFFECTS OF ADAPTIVE RESPONSES

It is a general rule of pharmacology that pharmacodynamic responses are often followed by adaptive responses which, crudely put, are the body's attempt to "overcome" or "counteract" the effects of the drug. An example is the increase in the concentration of the membrane-bound enzyme Na^+/K^+ ATPase that occurs during continued treatment with cardiac glycosides such as digoxin. As cardiac glycosides exert their effects by inhibiting Na^+/K^+ ATPase, the localized increase in the concentration, or up-regulation, of this enzyme that occurs during therapy may be responsible for the relatively transient inotropic effects of the drugs that are seen in some individuals.

Cell surface adrenoceptors tend to up-regulate during prolonged administration of β -adrenergic blocking agents such as propranolol, resulting in increased numbers of active β -receptors. If the beta-blocking drug is withdrawn rapidly from a patient, a large number of β -receptors become available to bind to norepinephrine and epinephrine, their natural ligands. This can produce tachycardia, hypertension, and worsening angina—the so-called "beta-blocker withdrawal syndrome."

In some cases, the mechanisms of apparent adaptive responses have not yet been fully explored. For example, among subjects taking nonsteroidal anti-inflammatory drugs (NSAIDs), endoscopic studies have documented gastrointestinal mucosal damage within days of commencing treatment. Endoscopic investigation of patients chronically exposed to aspirin found that the mucosal damage appeared to resolve over time. While this suggests that continued

exposure promotes gastric adaptation, the mechanism by which this might occur is unclear. Pharmacoepidemiology studies of gastric ulceration and its complications of bleeding, perforation, and stenosis were in keeping with the observation, suggesting that the risk of gastrointestinal complications was highest in the early weeks of NSAID treatment, and declined thereafter. However, more recent evidence has questioned this conclusion. In a record linkage study, MacDonald *et al.* (1997) found that the increased risks of admission to hospital with gastrointestinal complications related to NSAID use were constant during continuous exposure and that excess risk appeared to persist for at least a year after the last exposure. So the experimental and observational studies of adaptation are somewhat at odds in their findings, which illustrates that it is not always possible to correlate our biologic understanding with epidemiologic observations; sometimes the latter can inform the former, a reversal of a common view of the discovery process. In the case of NSAIDs, the field of study has tended to be overtaken by the introduction of the controversial COX-2 inhibitor drugs (e.g., rofecoxib), which have a lower risk of causing serious gastrointestinal complications than non-selective NSAIDs, but increase the risk of vascular occlusion.

EFFECTS OF AGE

On the whole, the effects of age on pharmacodynamic responses have been less well studied than its effects on pharmacokinetics. This is particularly so in the very young, who are rarely included in experimental studies to investigate the clinical effects of drugs. Although it may seem counterintuitive, the elderly are often *equally* or even *less* sensitive to the primary pharmacologic effects of some drugs than are the young. But, overall the elderly behave as though they have a “reduced functional reserve” and their secondary homeostatic responses may be impaired. Several examples of the effects of old age on pharmacodynamic responses may be found in the cardiovascular therapeutics:

- It has long been known that elderly subjects are relatively resistant to the effects of both the β -agonist drug isoproterenol and the β -blocking drug, propranolol. The extent to which this is due to elevated levels of plasma catecholamines and alterations in β -receptor numbers is not clear.
- Elegant experimental work has demonstrated that elderly subjects have a blunted primary electrophysiologic response to the calcium channel blocking drug verapamil. The degree of prolongation of the electrocardiographic

P–R interval in response to a given concentration of verapamil was *less* pronounced in elderly than in younger subjects. However, in contrast to its effect on the P–R interval, verapamil produces a *greater* drop in blood pressure in the elderly than it does in younger subjects. How may the last two observations be reconciled? The likely answer is that both the secondary adaptive physiologic responses and the primary pharmacologic response are impaired in the elderly subjects. Maintenance of blood pressure depends on activation of the sympathetic nervous system, which tends to be less responsive in the elderly. It is likely that impairment of secondary (adaptive) responses, rather than increased sensitivity to the primary pharmacologic actions *per se*, accounts for the increased susceptibility of elderly subjects to the side effects of many drugs.

Homeostatic regulation (the body’s control of its internal environment) is often impaired in the elderly and may contribute to the occurrence of adverse events as well as increased sensitivity to drug effects. For example, older individuals have an impaired ability to excrete a free water load, possibly as a result of lower renal prostaglandin production. This may be exacerbated by treatments that further impair either free water excretion, such as diuretics, or renal prostaglandin production, for example, NSAIDs. In either case, there is a risk of dilutional hyponatremia or volume overload.

Postural hypotension (the sudden drop in blood pressure that occurs with standing or sitting up, particularly in patients on antihypertensive drugs) is frequently symptomatic in the elderly; the pathogenesis probably includes decreased baroreceptor response, altered sympathetic activity and responsiveness, impaired arteriolar and venous vasomotor responses, and altered volume regulation. Accordingly, drugs that alter central nervous system function, sympathetic activity, vasomotor response, cardiac function, or volume regulation may exacerbate postural changes in blood pressure. The list of agents is extensive and includes such commonly used drugs as phenothiazines, antihypertensives, diuretics, and levodopa.

EFFECTS OF DISEASE STATES

The effects of disease states on pharmacodynamics have not been widely studied. As most diseases that lead to organ failure are more common in older subjects, the effects of disease can be confounded by age. It is a common clinical observation that individuals with certain diseases can have exaggerated responses to particular drugs. For example, individuals

with chronic liver or lung disease sometimes exhibit extreme sensitivity to drugs that depress central nervous system function, such as benzodiazepines and opiates. This apparent increase in drug sensitivity may be due to: (i) changes in receptor function, which would increase actual sensitivity to drugs, or (ii) disease-related changes in neuronal function, such as occurs in encephalopathy caused by severe lung or liver disease. A further possibility, in the case of liver failure, is the presence of elevated concentrations of circulating endogenous ligands that bind to the benzodiazepine receptor, the effect of which is additive to that of diazepam.

Another example of the role of disease states in pharmacodynamic variability is the propensity for NSAIDs to impair renal function in certain groups of individuals. Both congestive heart failure and hepatic failure are characterized by high circulating levels of the vasoconstrictor hormones norepinephrine, angiotensin II, and antidiuretic hormone. In response to the presence of these hormones, the kidneys release prostaglandins to modulate their vasoconstrictor effects and thus help preserve renal blood flow in times of physiologic stress. In susceptible individuals, inhibition of prostaglandin synthesis (e.g., as a result of NSAID administration) can lead to unopposed vasoconstriction with a marked and rapid reduction in renal blood flow, and a consequent fall in the rate of glomerular filtration.

DRUG-DRUG INTERACTIONS THAT OCCUR THROUGH PHARMACODYNAMIC MECHANISMS

Although many important drug-drug interactions occur through pharmacokinetic mechanisms, a number of important interactions are pharmacodynamic in nature. Pharmacodynamic interactions arise as a consequence of drugs acting on the same receptors, sites of action, or physiological systems and having either synergistic or antagonistic effects. In examining the variability that exists within the population with regard to the effects of drugs, the presence or absence of concomitant medications can play a particularly important role and must be considered as potential causal or confounding variables in pharmacoepidemiology studies.

For example, individuals with any given serum digoxin concentration are more likely to suffer from digoxin toxicity if they are depleted of certain electrolytes, such as magnesium and potassium. Thus, patients on concomitant magnesium-/potassium-wasting diuretics such as furosemide are more likely than those who are not to develop arrhythmias, given the same serum digoxin concentration.

Many drugs have central nervous system depressant effects and these may be potentiated where a number of such agents are used together, such as hypnotics, anxiolytics, antidepressants, opioids, anti-epileptics, antihistamines, and methyl dopa. A "serotonergic syndrome" (consisting of mental changes, muscle rigidity, hypertension, tremor, hyperreflexia, and diarrhea) may be induced in some patients given combinations of proserotonergic drugs such as selective serotonin reuptake inhibitors (SSRIs), tramadol, tricyclic antidepressants, monoamine oxidase inhibitors (MAOIs), carbamazepine, and lithium.

Competition between drugs acting at the same receptor sites usually results in antagonistic effects. These may be desired, as in the case of naloxone or flumazenil given to reverse central nervous system depression or coma resulting from opiate or benzodiazepine overdose, respectively, or unintended, as in the case of the mutual antagonism occurring between β -agonists (bronchodilators) and non-selective β -antagonists (bronchoconstrictors).

Sildenafil selectively inhibits cyclic guanosine monophosphate (cGMP)-specific phosphodiesterase type 5, the predominant enzyme metabolizing cGMP in the smooth muscles of the corpus cavernosum. By doing so, it restores the erectile response to sexual stimulation in men with erectile dysfunction. However, the formation in the first place of cavernosal cGMP is due to the release of nitric oxide in response to sexual stimulation. In men taking concomitant nitrate drugs for heart disease, there is a risk of a precipitous fall in blood pressure due to potentiation by sildenafil of their hypotensive effects, mediated by vascular smooth muscle relaxation. Although concomitant use is contraindicated by the manufacturer, a number of sildenafil-associated deaths are thought to have been due to this drug combination.

Case-control pharmacoepidemiology studies have demonstrated an association between long-term use (>3 months) of certain appetite suppressants (phentermine plus fenfluramine or dexfenfluramine, or dexfenfluramine alone) and cardiac valve abnormalities. The use of amphetamine-like appetite suppressants, mainly fenfluramine and dexfenfluramine, has also been associated with primary pulmonary hypertension. It has been postulated that these unintended effects were due to serotonin accumulation as a consequence of both increased release and reduced removal of serotonin. Serotonin is the predominant mediator of pulmonary vasoconstriction caused by aggregating platelets and has been shown to increase pulmonary vascular smooth muscle proliferation. Prolonged use of fenfluramine and dexfenfluramine may produce an excess of serotonin sufficient to damage blood vessels in the lungs. Serotonin

excess is also thought to be responsible for the cardiac damage as the pathological findings in damaged valves resembled those of carcinoid heart disease or heart disease associated with ergotamine toxicity, both of which are serotonin-related syndromes. Both fenfluramine and dexfenfluramine were withdrawn from the worldwide market in 1997.

In conclusion, adaptive responses, age, disease states, and concomitant medications can each have important effects on pharmacodynamic responses, and may result in considerable heterogeneity in the responses to drugs, both between and within individuals. Allowance must be made for this when interpreting pharmacoepidemiologic data.

We will now consider the effects of pharmacokinetic determinants of variability in drug response.

THE ROLE OF PHARMACOKINETICS IN DETERMINING VARIABILITY OF DRUG RESPONSE

As noted earlier, pharmacokinetics is the science that describes the time course of the absorption, distribution, and elimination of drugs within the body, the processes which in turn determine the concentration of drug at its active site. From a research perspective, it is generally easier to measure changing concentrations of drugs in body fluids than it is to characterize the pharmacologic responses to those concentrations. Consequently, the literature on pharmacokinetics is voluminous, and it could be said that clinical pharmacology as a discipline has been overly concerned with its study. However, it must be acknowledged that variation in pharmacokinetic parameters is an important cause of the observed heterogeneity that exists with regard to patients' response to drugs. In this section, we review the processes of absorption, distribution, and elimination of drugs, and then consider the effects of age, genetics, disease, and concomitant medications. First, however, it is useful to define some of the basic mathematical parameters that are used in pharmacokinetics.

BASIC MATHEMATICAL PARAMETERS USED IN PHARMACOKINETICS

Figure 4.1 shows the serum concentration of a hypothetical drug following a single intravenous bolus injection, plotted against time. Because the rate of decline of serum drug concentrations, like many other natural phenomena, frequently appears to be log-linear, the vertical axis is plotted on a logarithmic scale. It was observed that, for some drugs, the initial portion of the log-concentration versus time

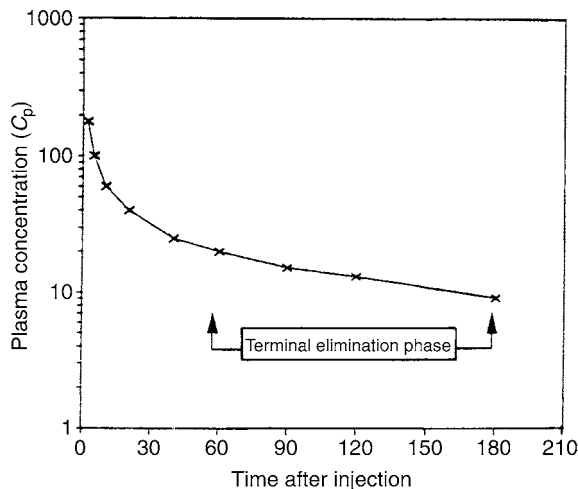


Figure 4.1. Plasma concentration/time curve of a hypothetical drug after intravenous administration. Note the early rapid decline in blood levels which reflects distribution of the drug, a consequence of its lipid solubility and degree of protein binding. The terminal phase of the concentration/time curve is log-linear and reflects elimination of the drug from the central compartment. This may be by renal clearance or hepatic metabolism. In this case, the terminal elimination phase is equivalent to a half-life of about 100 minutes.

curve deviates notably from the line that is defined by the terminal portion of the curve. For this reason, the concept of pharmacokinetic compartments was developed. A *pharmacokinetic compartment* is a theoretical space, defined mathematically, 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. Because of the rather theoretical basis of mathematical modeling clinical pharmacologists have, more recently, been trying to correlate plasma concentration time curves more closely with physiological parameters such as cardiac output and tissue partition coefficients. The incorporation of these variables sometimes improves the accuracy and predictive ability of kinetic models.

The initial, rapid decline of measured drug concentration (Figure 4.1) is attributed to distribution of drug molecules through plasma and into other well-perfused tissues. This is usually referred to as the *distribution phase*. After the concentration of drug molecules has reached equilibrium across the compartments the more gradual decline in serum concentrations that is seen at the right-hand portion of the

curve represents the elimination of drug from the body, and is referred to as the *terminal elimination phase*.

Because the dose of injected drug is known, and the initial plasma concentration immediately following administration (C_p , the peak plasma concentration) can be extrapolated from the points on the curve, a pharmacokinetic parameter known as *apparent volume of distribution*, or V_d , can be calculated by dividing dose by C_p ; V_d is expressed in units of volume, such as liters, and is the volume into which the drug *appears* to have been dissolved in order to produce the actual peak concentration. Just as with pharmacokinetic compartments, the apparent volume of distribution is a theoretical, rather than an actual, volume, although it does have some physiologic interpretability. For example, a highly lipid-soluble drug such as a tricyclic antidepressant may have an apparent volume of distribution of hundreds of liters. This is because the drug partitions readily into fatty tissue, leaving little measurable drug in the bloodstream.

The slope of the line that represents the elimination phase is known as the *elimination rate constant*, or K_e , and is expressed in units of reciprocal time, such as hour^{-1} . Because of the linearity of the terminal elimination phase on a semi-log plot, the time that it takes for any given drug concentration to decline to half of its original concentration is constant, and is known as the drug's *half-life*, or $T_{1/2}$. Half-life is expressed in units of time, such as hours. Mathematically, this parameter is calculated from the elimination rate constant using the formula: $T_{1/2} = 0.693/K_e$.

An additional pharmacokinetic parameter, *clearance*, or Cl , can also be defined. Again, this is a theoretical parameter, and refers to the volume whose concentration is being measured that *appears* to be completely cleared of drug, per unit of time, regardless of the clearance mechanism. It is expressed in units of volume per unit time, such as liters per hour. Clearance can be calculated by taking the product of the apparent volume of distribution and the elimination rate constant. It is important to note that both volume of distribution and the rate of clearance together determine the half-life of elimination ($T_{1/2} = (0.693 \times V_d)/Cl$).

When a drug is administered according to a stable dosing regimen, the plasma drug concentration eventually reaches an equilibrium state, in which the amount of drug being administered equals the amount of drug being eliminated from the body (Figure 4.2). This is referred to as *steady state*. The amount of time required for a drug to reach steady state depends on the rate of elimination of the drug from that individual. For the purposes of therapeutic drug monitoring, achieving >95% of the steady state concentration is generally considered sufficient in order to estimate the true steady state concentration. This can be accomplished by

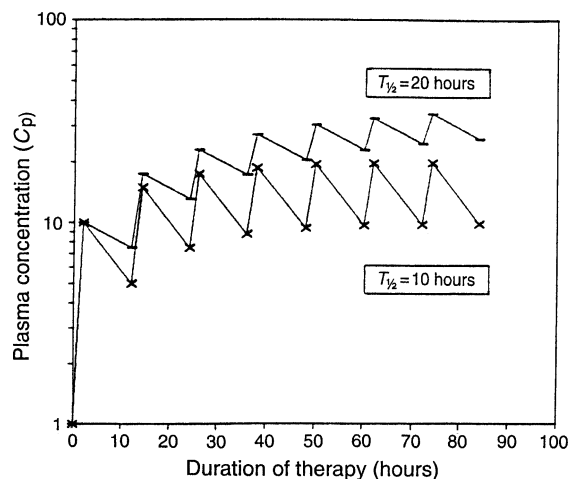


Figure 4.2. Profile of plasma drug concentrations for two hypothetical drugs during a repeated 12-hourly dosing schedule. The lower curve relates to a drug with a half-life of 10 hours; steady state concentrations are achieved after 50–60 hours. In contrast, the upper curve relates to a drug with a half-life of 20 hours and the plasma concentrations are still rising at the end of the study.

obtaining a biologic sample after approximately five drug half-lives. The amount of drug that is administered affects the *magnitude* of the steady state drug concentration, but not the *amount of time* that it takes to reach steady state. Under these circumstances a longer half-life results in a longer time to achieve steady state concentrations and a tendency to accumulate. This can lead to toxicity when a long half-life drug is commenced in hospital and suitable arrangements are not made for follow-up and monitoring of either drug concentrations or effects.

EFFECTS OF VARIATIONS IN DRUG ABSORPTION

In clinical practice, most drugs are administered by mouth. Because most drug molecules are small and at least partially lipid soluble, they are absorbed by *passive diffusion* across the large surface area of the mucosa lining the small intestine. The extent of absorption is determined primarily by the physicochemical properties of the drug and the integrity of the small intestine, while the *rate* of absorption depends largely on gastric emptying time and the motility of the small intestine. Both increasing age and the presence of disease states sometimes affect the extent and rate of gastrointestinal absorption of drugs; if they have an effect,

both will tend to *decrease* absorption. Reduced absorption can also result from co-ingestion of drugs with chelating agents. For instance, the anion-binding resin cholestyramine (used to bind bile acids and reduce blood cholesterol levels) is capable of binding a variety of drugs, including statins (e.g., fluvastatin, simvastatin), which may be co-ingested by patients with severe hypercholesterolemia.

Rarely, variations in absorption may increase the rate or extent of the systemic availability of a drug and cause adverse effects. This is usually explained by altered bioavailability following a formulation change. An example of this was seen in Australia involving the calcium antagonist nifedipine, which was available both as sustained release tablets and as rapid release capsules. Individuals who were switched inadvertently from the former to the same dose of the latter sometimes experienced hypotension, presumably due to rapid absorption leading to elevated peak concentrations of the drug, with subsequent vasodilatation.

Absorption of drugs is not confined to the gastrointestinal tract. Systemic absorption of drugs may occur following unintended absorption via other routes, such as transdermally, following administration by metered-dose inhaler, or ocular instillation. Each of these may result in adverse drug effects. The ability of lipid-soluble compounds to be absorbed across intact skin has been utilized in the design of transdermal delivery systems for several drugs, including estradiol, nitroglycerin, nicotine, and scopolamine.

Transdermal drug absorption can produce adverse as well as beneficial effects, as illustrated by the hexachlorophene toxicity that occurred in neonates following the mixture of excessive quantities of this antiseptic with talcum powder, and in another instance following the inadvertent contamination of talcum powder with the anticoagulant warfarin. Neonates are particularly susceptible to the effects of transdermal drug exposure because their skin provides a poor barrier to systemic absorption, and because they have a large surface area in proportion to their body weight. In a similar fashion, quantities of corticosteroid sufficient to produce systemic effects can be swallowed following administration from a metered-dose inhaler. Beta-blocking drugs instilled into the eyes can travel down the naso-lachrymal duct to be swallowed and absorbed, inducing bronchospasm or exacerbation of congestive heart failure in susceptible individuals.

In summary, because variability in the absorption of oral dosage forms of drugs from the gastrointestinal tract typically *reduces* absorption, it is more important as a cause of lack of efficacy than an increase in adverse effects. However, unintended systemic absorption can occur through a variety of routes, and can have important consequences.

EFFECTS OF VARIATION IN SYSTEMIC DISTRIBUTION OF DRUGS

As drug molecules are absorbed, they are distributed to various tissues at a rate and to an extent that are determined by: (i) the lipid solubility of the drug, (ii) the degree of protein binding of the drug, and (iii) the amount of blood flow received by the different tissues. A high degree of lipid solubility confers an ability to move readily across cell membranes and to accumulate in lipid environments, and therefore results in a higher proportion of drug molecules being distributed to fatty tissues. Extensive binding to plasma proteins will reduce movement of drug molecules out of the central compartment, and thus reduce the drug's apparent volume of distribution. Better perfused tissues will tend to receive a larger amount of drug than tissues which are poorly perfused.

Protein binding is an aspect of drug distribution that receives considerable attention—perhaps more than it deserves. Untoward effects of drugs are often attributed to altered protein binding, which can occur in certain disease states, pregnancy, or when other highly protein-bound drugs are taken concurrently. However, there are relatively few occasions when disease-induced disturbances of protein binding or protein binding drug-drug interactions have been shown to have clinically important effects. The main reason for this is that, while it is the free fraction of drug that interacts with target proteins in order to produce a pharmacologic effect, it is also the free fraction that is available to the clearance mechanisms. Therefore, any increase in the free drug concentration that is caused by either reduced albumin levels or by displacement by other drugs is also accompanied by an increase in clearance, so that the free, active concentration ultimately changes little. The period of time for which there may be an important increase in free concentration is limited to about three half-lives after onset of the interaction. There are occasional situations in which this general rule may not apply, particularly where small changes in concentrations have large effects, or where the drug has a single clearance mechanism that has a limited capacity and therefore can become saturated.

Research has revealed that not all drug distribution is passive. There is growing interest in the function of drug “transporters.” Drug transporters are specialized proteins that mediate the efflux of drugs (i.e., transport out) from cells and tissues. The most studied mediator is P-glycoprotein, which is located in the plasma membrane, and translocates its substrates from the inside of the cells to the outside. Interest in this protein arose from the observation that over-expression of P-glycoprotein in cancer cells leads to low intracellular concentrations of anti-cancer drugs

and apparent resistance of the tumors to treatment. Subsequent research has revealed a wider potential for this transporter to explain variations in the actions of a range of compounds, by reducing concentrations in various tissues, including the central nervous system. For instance, recent work has documented polymorphism of the drug transporter gene ABCB1, which codes for P-glycoprotein, and causes resistance to a wide range of anticonvulsant drugs.

In conclusion, changes in drug distribution are often cited as reasons for variability in response to drugs and, therefore, may be implicated when seeking explanations for pharmacoepidemiologic findings. Alterations in the passive phases of drug distribution rarely produce clinically important effects. However, new information on the role of drug transporters suggests that variability in these active processes may account for lack of efficacy to various drug classes.

EFFECTS OF VARIATION IN DRUG ELIMINATION

Drugs are excreted from the body either as the unchanged parent compound, or as one or more products of drug metabolism. Although a number of organs, including the biliary system, the lungs, and the skin, participate in drug elimination, the kidneys play the most important role. Most excretory organs remove water-soluble compounds more efficiently than they remove lipid-soluble compounds. Consequently, water-soluble drugs tend to be eliminated unchanged in the urine, while lipid-soluble drugs tend to undergo metabolism to more water-soluble products, usually in the liver, prior to being excreted.

Effects of Variation in Renal Elimination

Virtually all drugs are small enough to be filtered through the glomeruli, the filtering units of the kidney, into the renal tubules. The extent of *glomerular filtration* depends on both the perfusion pressure to the glomeruli, and the protein binding characteristics of the drug. Because only the unbound fraction of a drug in the bloodstream is available to be filtered, a high degree of protein binding and a high affinity of the drug for the binding protein will limit the amount of drug that reaches the renal tubules. Once inside the renal tubules, lipid-soluble drugs are readily reabsorbed into the bloodstream, across the lipid membranes of the cells lining the renal tubules, leaving virtually none of the filtered fraction to be excreted in the urine. Because this process does not involve the consumption of cellular energy, it is known as *passive tubular reabsorption*. Water-soluble drugs, such as aminoglycoside antibiotics and digoxin, do not cross the

tubular membrane and therefore remain in the urine and are excreted. *Active tubular secretion* occurs when substances are secreted into the renal tubules by energy-consuming carrier proteins. It is an important clearance mechanism for a number of drugs, including penicillin.

For drugs that are readily ionized at physiologic pH, such as salicylates, pH can be a crucial determinant of renal excretion. Because the non-ionized (uncharged) drug fraction is the most lipid soluble, it is most likely to undergo passive tubular reabsorption. Therefore, the renal excretion of salicylates, which are ionized at high (alkaline) pH, can be enhanced by the pharmacologic alkalization of the urine. This characteristic is exploited when alkaline diuresis is used to enhance renal clearance in cases of salicylate poisoning.

From a pharmacoepidemiology standpoint, the importance of renal clearance is that it can be estimated and, therefore, individuals can be identified who are at risk of toxicity through accumulation of water-soluble drugs. This is much simpler than estimating hepatic function (see below).

Plasma creatinine concentration is a measure of renal function that is frequently used in clinical practice. The rate at which the kidneys clear creatinine from the blood (creatinine clearance) correlates closely with the glomerular filtration rate. Creatinine concentration at any point in time is a function of production and clearance, both of which tend to decline to a proportionally similar degree with age—the former because of declining muscle mass, the latter because of an age-related decline in numbers of functioning glomeruli. For example, a blood creatinine level of 0.1 mmol l^{-1} in an 80-year-old female reflects a much lower level of renal function than the same creatinine concentration in a 20-year-old male (Figure 4.3).

The importance of considering age when interpreting plasma creatinine concentrations is illustrated in Figure 4.3. If both subjects mentioned in the previous paragraph required treatment with digoxin, the dose used to achieve therapeutic concentrations in the older subject would be less than half that required by the young man. Remember that these individuals have identical blood creatinine concentrations, illustrating the limitations of relying on this parameter solely as a measure of renal function.

In conclusion, it is important to take account of variation in renal function when conducting pharmacoepidemiology studies of drugs for which this is the principal route of elimination from the body. Some pharmacoepidemiology studies access clinical laboratory data, enabling estimates of renal function to be made and included in analyses of outcomes. Consequently, it is important to recognize that plasma creatinine concentrations must be adjusted for age and body

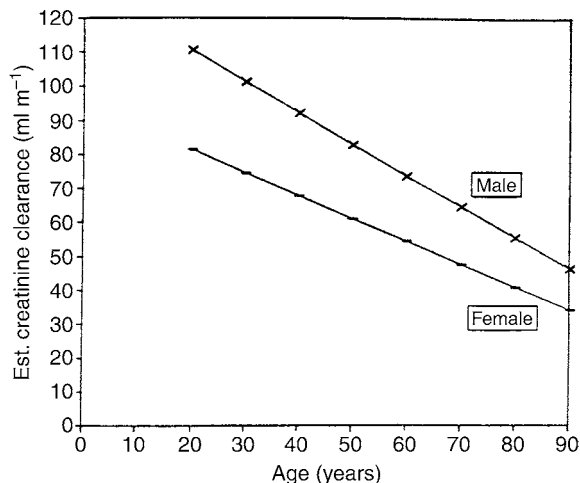


Figure 4.3. Change in estimated creatinine clearance (Cockcroft and Gault Formula 7) with age in a male and a female who maintain a serum creatinine of 0.1 mmol l^{-1} (NR $0.07\text{--}0.12 \text{ mmol l}^{-1}$) throughout their lives. In estimating creatinine clearance, it was assumed that the male maintained a weight of 75 kg, and the female a weight of 60 kg. The figure indicates that creatinine clearance declines in a linear fashion with age, and serum creatinine, alone, is an inadequate measure of glomerular filtration. Consequently, the clearance of some drugs is impaired in the elderly. For instance, the female depicted at 80 years (creatinine clearance 45 ml min^{-1}) would require less than half the dose of digoxin taken by the male at 20 years, despite having an identical serum creatinine level.

weight before being used as an estimate of renal function. A number of suitable formulae have been published, with the most widely used being that of Cockcroft and Gault (1976).

Drug–Drug Interactions Involving Renal Elimination

Drugs are capable of interfering with elimination of other substances by the kidney. This can occur through an effect on filtration, tubular reabsorption, or tubular secretion. A thorough discussion of this topic is beyond the scope of this overview, but one or two examples will illustrate the importance of this type of interaction. The deleterious effect of NSAIDs on renal blood flow that occurs in certain clinical states was mentioned earlier. As a result, NSAIDs are capable of inhibiting the clearance of a range of potentially toxic compounds, including lithium and methotrexate. Accumulation of these agents can produce serious adverse effects.

In cases where filtration pressure is maintained by angiotensin II-mediated vasoconstriction of the post-

glomerular efferent arteriole, angiotensin converting enzyme inhibitors (ACEIs) or angiotensin receptor antagonists may abruptly decrease the glomerular filtration rate through inhibition of angiotensin II synthesis. This may occur in renal artery stenosis, hypovolemia, and cardiac failure, thereby increasing the effects or the toxicity of concomitantly administered drugs that are renally excreted or that are nephrotoxic. The immunosuppressant cyclosporine induces vasoconstriction of the afferent glomerular arteriole in a dose-related and reversible fashion. An increased risk of acute renal failure exists when cyclosporine is combined with NSAIDs, ACEIs, or other nephrotoxic drugs.

Probenecid, a drug that is used in the treatment of gout, reduces the reabsorption of uric acid by the renal tubules, and inhibits the active tubular secretion of penicillin. These actions explain two therapeutic effects of probenecid—it lowers uric acid concentrations in the blood and enhances the effect of a dose of penicillin. Both mechanisms are exploited in clinical practice.

EFFECTS OF VARIATION IN DRUG METABOLISM

Variability in the metabolism of drugs is an important factor to be considered in the analysis and interpretation of pharmacoepidemiology studies. In this section, we will consider the effects of genetics, age, disease states, and concomitant drugs on the metabolism of drugs. Next, we will discuss some of the implications of active drug metabolites and intrinsic clearance. But first, an overview of drug metabolism is in order.

An Overview of Drug Metabolism

The majority of drugs are too lipid soluble to be effectively eliminated by the kidneys. First, they must be converted into water-soluble metabolites that can then be excreted in the urine, or sometimes in feces, via the bile. The metabolic steps necessary for the conversions occur primarily in the liver. Chemical reactions that result in the metabolism of drugs are classified as either Phase I or II reactions. *Phase I reactions* are usually oxidative (e.g., hydroxylation) and create an active site on the drug molecule that can act as a target for Phase II conjugative (synthetic) reactions. *Phase II reactions* involve the synthesis of a new molecule from the combination of the drug and a water-soluble substrate such as glucuronic or acetic acid (Figure 4.4). The product of this type of reaction, for instance the glucuronide or acetyl derivative of the drug, is highly water soluble, and is

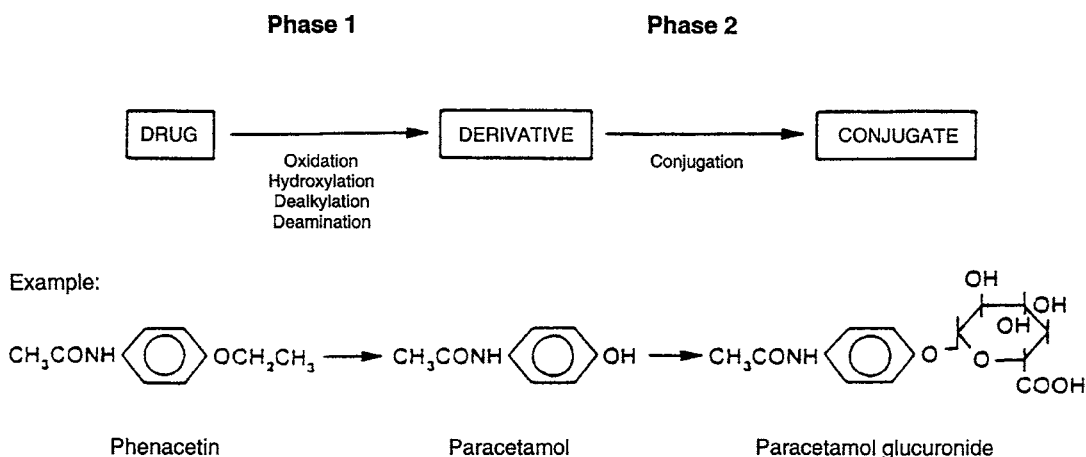


Figure 4.4. Phase I and II reactions often occur sequentially. Phase I reactions usually consist of oxidation, reduction, hydrolysis, and products are often more reactive, and sometimes more toxic, than the parent drug. Phase II reactions involve conjugation and these usually result in inactive compounds. The main effect of this conjugation is to render the substance more water soluble. In the example given, phenacetin is converted to acetaminophen (paracetamol) by dealkylation (Phase I reaction). This introduces a reactive hydroxyl group to which the glucuronyl group can be attached. Both phenacetin and acetaminophen are active, whereas acetaminophen glucuronide is inactive, and water soluble, and is excreted in the urine. (Figure and legend reproduced with permission from Rang *et al. Pharmacology*, 5th edn. Edinburgh: Churchill Livingstone, 2003.)

excreted in the urine, or occasionally in the feces, if it is of high molecular weight.

Most drugs that undergo Phase I (oxidative) metabolism are transformed by a superfamily of enzymes called the cytochrome P450 (CYP) system. Cytochrome P450 is so named because in a certain form its maximal light absorption occurs at a wavelength of approximately 450 nanometers. Most Phase I drug metabolism involves cytochrome P450 families 1, 2, and 3 (CYP1, CYP2, and CYP3). Specific enzymes exist within CYP families. For example, enzymes CYP2C9, CYP2C10, CYP2C18, and CYP2C19 are responsible for most drug metabolism within the CYP2C group of enzymes. Different drugs may be metabolized by different isoenzymes, or because of incomplete substrate specificity a given drug may be metabolized by more than one enzyme.

Some drugs are capable of participating in synthetic reactions without prior Phase I metabolism. An example is the benzodiazepine temazepam, which is conjugated directly with glucuronide, and is eliminated in the urine in this form. In contrast, diazepam, another benzodiazepine, must undergo several Phase I oxidative reactions before it can be conjugated and eliminated. Phase I reactions are usually the rate limiting step in this process and are subject to much greater intra- and inter-individual variability than are Phase II reactions. This explains why diazepam metabolism

is largely affected by age and disease, while temazepam metabolism is relatively unaffected by these factors.

Effect of Genetic Factors on Drug Metabolism

Genetic factors are sometimes important in determining the activity of drug metabolizing enzymes. Studies have shown that half-lives of phenylbutazone and coumarin anti-coagulants are much less variable in monozygotic than in heterozygotic twins. The half-lives of these drugs in the overall population display an approximately Gaussian distribution, although the limits are often wide, and may encompass 5- to 10-fold variations.

The metabolism of the anti-tubercular drug isoniazid exhibits a bimodal distribution within the population. The conjugation of isoniazid with acetic acid is an important step in its inactivation and elimination. Variability in the rate of isoniazid acetylation results from a single recessive gene whose distribution shows some racial dependence (acetylation polymorphism). For example, approximately half (50–60%) of most Caucasian communities are slow acetylators, and therefore have a reduced capacity to eliminate the drug. In Japan, the prevalence of the slow acetylator phenotype is only 15%, and slow acetylators have not been identified in Eskimo populations. Although attempts have been made to correlate acetylator phenotype with risk

of isoniazid-induced hepatotoxicity, published reports are equivocal, with some showing an association with slow inactivators and others showing an association with rapid inactivators. Recent work has emphasized the possible role of an alternative pathway for isoniazid metabolism. Patients with a particular CYP2E1 genotype had a higher risk of hepatotoxicity with isoniazid after adjustment for acetylator status.

Acetylation polymorphism affects the metabolism of a number of drugs in addition to isoniazid; these include some sulphonamides (including sulphasalazine), hydralazine, procainamide, dapsone, nitrazepam, and caffeine. In general, the clinical implications are that slow acetylators require lower doses both for therapeutic effect and to minimize toxicity and side effects.

Hydroxylation polymorphism was identified in 1977. It has since been established that around 10% of Caucasians and 1% of Asians exhibit hydroxylation deficiency as a result of reduced activity of the enzyme CYP2D6. First described in relation to debrisoquine, the deficiency also affects the metabolism of antidepressants (amitriptyline, clomipramine, desipramine, nortriptyline, mianserin, paroxetine), antiarrhythmics (flecainide, propafenone), antipsychotics (haloperidol, perphenazine, thioridazine), and β -blockers (alprenolol, metoprolol) leading to accumulation of the active parent compound. In the cases of amitriptyline and thioridazine, both parent and active metabolite accumulate. Poor hydroxylators may have markedly increased effects or a prolonged duration of action of the affected drugs.

CYP2C19 polymorphism is described in 2–5% of Caucasians and 12–23% of Asians who have a deficient capacity to hydroxylate S-mephenytoin. CYP2C19 also catalyzes the metabolism of commonly used drugs such as barbiturates, omeprazole, propranolol, diazepam, and citalopram.

The clinical consequences of genetic polymorphism have not been fully elucidated, but it is likely that such genetically determined differences may account in some part for the inter-individual and inter-ethnic differences in therapeutic response and side effect profile observed with many drugs.

The CYP2D6 phenotype of a given individual can be determined by testing the metabolic clearance of a test drug, such as debrisoquine or sparteine. This technique can be useful in performing pharmacoepidemiology studies. For instance, Wiholm *et al.* (1981) compared debrisoquine hydroxylation in a group of subjects who had developed lactic acidosis while taking phenformin with the expected distribution in the Swedish population. This study illustrates the potential for investigating groups of individuals who

display apparently idiosyncratic reactions to certain drugs. Other examples of the use of laboratory techniques to investigate the occurrence of serious adverse reactions include the demonstration of possible familial predispositions to halothane hepatitis and phenytoin-induced hypersensitivity syndromes.

Genetic polymorphism of drug metabolizing enzymes does not just account for adverse effects, it may also lead to lack of efficacy. For instance, individuals who have the extensive metabolism genotype of CYP2C19 need large doses of proton pump inhibitors (e.g., omeprazole) to reduce gastric acid secretion. The slow metabolizer genotype of CYP2D6 will have poor conversion of codeine to morphine and consequently will not experience optimal analgesic effects of the drug.

The large series of well-validated case reports held by many spontaneous reporting systems represent fertile areas for this type of research (see Chapters 7 and 8). However, this requires that the adverse reactions agency maintains a good relationship with those who send in reports. The agency also must have a mechanism for obtaining access to biological or genetic material with the approval of the relevant ethics committees. The use of genetic testing in concert with voluntary adverse drug reaction (ADR) reports and pharmacoepidemiologic methods to predict and explain variability in drug response is a promising new area of research (see also Chapter 18).

Effects of Disease on Drug Metabolism

Hepatic disease can result in reduced elimination of lipid-soluble drugs that are metabolized by this organ. Unfortunately, there are no convenient tests of liver function that are analogous to the measurement of creatinine clearance for estimating renal function. The conventional biochemical tests largely reflect liver damage, rather than liver function. It is quite possible for an individual to have grossly disordered liver function tests, while still metabolizing drugs normally, or alternatively to have apparently normal liver function tests, despite the presence of advanced liver disease with marked impairment of metabolic capacity. To complicate matters further, the liver behaves as though it has a number of “partial functions” that respond differently to disease. For example, bilirubin conjugation may be impaired, while albumin synthesis continues fairly normally. Alternatively, both of these functions may be almost normal, despite the presence of liver disease that has progressed so far that it has resulted in elevated pressure in the portal vein, with subsequent bleeding esophageal varices. It is thus difficult to generalize about the effects of

liver disease on hepatic drug metabolism. However, pharmacokinetic studies have shown that liver disease has to be severe, and usually chronic, to result in marked impairment of drug elimination. This is the case, for example, in individuals with cirrhosis or chronic active hepatitis, where Phase I reactions are primarily affected, while conjugative reactions are relatively spared. Other individuals, such as those with biliary obstruction or acute viral hepatitis, may have surprisingly normal drug metabolism.

Drug metabolism may also be affected by disease processes originating in other organs. For example, congestive heart failure can result in severe congestion of the liver, and therefore impair the hepatic clearance of some compounds, while hypoxia has been shown to reduce markedly the metabolism of theophylline. Reduced liver blood flow may also result in reduced extraction and metabolism of high clearance drugs such as morphine and propranolol.

To summarize, liver disease is a relatively uncommon cause of clinically important impaired drug metabolism. Generally, it can be stated that genetic and environmental factors are more important causes of variability in hepatic metabolism of drugs than diseases of the organ itself.

Effects of Active Metabolites

The general rule that drug metabolism produces metabolites that are inactive or markedly less active than the parent drug does not always hold true. This should be considered as a possible explanation for unexpected pharmacoepidemiologic findings. For example, several metabolites of carbamazepine contribute to its pharmacologic activity. The hydroxyl metabolite of propranolol has similar activity to its parent compound. Conjugated metabolites are usually devoid of activity, but morphine-6-glucuronide has been shown to have morphine-like action, and accumulation of this metabolite may explain the prolonged opiate effect of morphine that is found in individuals with advanced renal failure. Likewise, the conjugated acetyl derivative of the antiarrhythmic drug procainamide has been shown to have pharmacologic activity, and may cause toxicity.

Sometimes a metabolite has toxic effects that are not shown by the parent drug. *N*-Acetylbenzo-quinoneimine is the toxic metabolite formed by the oxidative metabolism of acetaminophen. This is normally produced in small quantities but rapidly cleared by reaction with glutathione. In acetaminophen poisoning, the available glutathione reserves are exhausted and the toxic metabolite is free to exert its action on cell membranes, leading to hepatic damage that may on occasions be fatal. More of the metabolite may be

formed in the presence of enzyme induction. As a result, chronic heavy drinkers and individuals taking long-term anticonvulsants may be more prone to develop liver damage.

Effects of Presystemic Clearance

Certain orally administered drugs are metabolized substantially in the intestine and/or in the liver before they ever reach the systemic circulation. This phenomenon is known as “first pass” metabolism or “presystemic” clearance. Drugs with high presystemic clearance include morphine, oral contraceptives, prazosin, propranolol, and verapamil. The differences between drugs with high or low presystemic clearances become apparent if hepatic metabolism is impaired by disease or inhibited by another drug or if blood flow to the liver is reduced in shock or in congestive heart failure.

In the case of a drug with low presystemic clearance, a reduction in hepatic metabolism results in a prolongation of the elimination half-life. Generally, it takes approximately five half-lives to reach a new steady state concentration, and accumulation of the drug may cause toxicity. If the drug has a high presystemic clearance, a decline in metabolism will result in increased *bioavailability* of the drug, with elevated, and possibly toxic, concentrations early in the course of treatment, possibly after the first dose, although it will still take five half-lives to reach the new steady state concentration (in a similar fashion as when the dose is increased). Thus, in a study of the adverse effects of drugs in subjects with hepatic impairment or metabolic inhibition by other drugs, the time course of adverse effects can be critically dependent on this factor.

Drug–Drug Interactions Involving Drug Metabolizing Enzymes

Enzyme induction occurs when the chronic administration of a substance results in an increase in the amount of a particular metabolizing enzyme. When such enzymes are induced, the rate of metabolism of a drug can increase several-fold. The subsequent fall in the concentration of the drug in the blood, and, consequently, at its sites of action, may result in a substantial loss of drug activity. For instance, failure of ethinylestradiol-containing oral contraceptives can result from the CYP450 enzyme-inducing effects of some anti-epileptic medications. The rate of metabolism of warfarin is increased by concomitantly administered drugs, including carbamazepine, rifampicin, and barbiturates, leading to reduced steady state plasma concentrations, and therefore a reduced anticoagulant effect.

Enzyme induction proceeds through a mechanism that involves increases in gene transcription, resulting in increased synthesis of new enzyme protein. It can take several weeks to reach its peak (except with alcohol, where the process is quicker), and can persist for some time after the inducing drug is ceased. CYP450 enzymes differ in their ability to be induced in response to a given exposure. For example, theophylline metabolism is readily inducible by cigarette smoking, while phenytoin metabolism is affected to a greater extent by barbiturates and anti-epileptic medications.

Enzyme inhibition occurs when the presence of one substance inhibits the metabolism of another substance. It involves either competition for active sites on the enzyme, or other binding-site interactions that alter the activity of the enzyme. In contrast to induction, enzyme inhibition occurs rapidly, and is rapidly reversed once the inhibiting substance has been withdrawn. As with induction, interacting compounds display considerable specificity, and a number of commonly used drugs have the capacity to inhibit microsomal function. For example, cimetidine is capable of inhibiting the metabolism of many compounds, including warfarin, theophylline, phenytoin, propranolol, and several benzodiazepines. In contrast, omeprazole has been shown to inhibit the metabolism of diazepam and phenytoin, but not of propranolol. Erythromycin is a clinically important enzyme inhibitor, well known for its effects on theophylline metabolism. Erythromycin and other macrolide antibiotics have been shown to inhibit the metabolism of the antihistamines terfenadine and astemizole and the prokinetic agent cisapride. These drugs may inhibit the potassium ion channels in the heart with a consequent risk of serious ventricular dysrhythmia, and they have been withdrawn from most markets. Drug–drug interactions are not always harmful. The calcium antagonists diltiazem and verapamil (but not nifedipine) increase cyclosporine plasma concentrations, but with relative sparing of nephrotoxicity, and the interaction has been used in clinical practice to produce an immunosuppressive concentration of cyclosporine at a lower ingested dose. Drug cost savings of 14–48%, attributable to the use of calcium antagonists, have been reported in transplant pharmacotherapy. Similarly, the protease inhibitor ritonavir is used in combination with other drugs in this class (e.g., lopinavir) in low doses as it inhibits their metabolism and “boosts” their blood levels and efficacy.

Recent research has revealed that many inducers and inhibitors of CYP3A4 act similarly on the drug transporter P-glycoprotein. For instance this pumps some drugs (e.g., digoxin) into the intestinal lumen, reducing bioavailability. Macrolide antibiotics may inhibit this transporter and so

increase the bioavailability of digoxin, leading to toxicity. Clearly the world of drug–drug interactions is more complex than we could have ever imagined!

Interactions arise not only as a consequence of other drugs; food constituents may affect drug metabolism. For example, the biflavonoids present in grapefruit juice have a strong inhibitory effect on the presystemic metabolism of calcium antagonists, causing a two- to three-fold increase in the systemic absorption of oral nifedipine and felodipine. A similar effect of biflavonoids on cyclosporine concentrations has been observed and has been utilized to reduce the doses, and therefore the side effects and costs, of cyclosporine therapy. Both calcium antagonists and cyclosporine are metabolized by the CYP450 isoenzyme CYP3A4, which is present in the gut wall and in the liver, and the biflavonoids inhibit its activity.

Sometimes dietary constituents can directly antagonize the effects of drugs. For instance, vitamin K-containing foods such as cabbage, brussels sprouts, broccoli, spinach, lettuce, rape seed oil, and soya bean oil, taken in sufficient amounts, may antagonize the effects of warfarin.

CONSEQUENCES OF VARIABILITY IN PHARMACOKINETICS

The foregoing discussion on causes of variability in pharmacokinetics is only of importance if there are clinical consequences that are likely to be detected in pharmacoepidemiology studies. Therefore, it is important to determine the circumstances in which these factors will contribute to variability in drug response.

Several factors are important. The first is the relationship between the concentration of the drug and its effects. Alterations of drug pharmacokinetics tend to be important if they involve drugs that have a low therapeutic ratio. This refers to the ratio of the concentration of drug that produces toxic effects to the concentration that elicits a therapeutic effect. If the ratio is low then small changes in drug concentration will lead to adverse effects. Examples of drugs with this profile are digoxin and lithium, which are primarily excreted unchanged by the kidneys, and theophylline and warfarin, which are primarily inactivated by hepatic metabolism. Cyclosporine also has a narrow therapeutic ratio, but wide variations between individuals in absorption, distribution, and metabolism have made definition of therapeutic, but nontoxic, concentrations difficult. It undergoes both hepatic metabolism and local metabolism in the gut, and the latter may be a major contributor to the variability in absorption.

Regardless of whether we are dealing with a decline in renal function or a reduction or inhibition of hepatic metabolism, the consequences in each case of increases in plasma concentration will be accumulation of the drug, and potential toxicity. In contrast, interactions involving drugs with high therapeutic ratios, for instance penicillin, will rarely produce significant adverse effects.

THE IMPORTANCE OF THE HUMAN FACTOR: PRESCRIBER AND CONSUMER BEHAVIOR

Human behavior may be an even greater source of variability in patterns of drug exposure than any other factor considered so far in this chapter. This is because many of the other factors considered in this chapter did not relate to use but rather to intensity of exposure (e.g., alterations in clearance, or drug–drug interactions). In contrast, non-adherence with therapy will have a more profound effect (see also Chapter 25).

In conducting pharmacoepidemiology studies, it is important to give cognizance to the impact of human behavior upon observed prescribing and consumption patterns. The influences that determine prescribing practices and consumer behavior are complex and have not been studied comprehensively; they are known to include factors related to the illness itself, the doctor, the patient, the doctor–patient interaction, drug costs and availability, perceived and actual benefits and risks of treatment, and pharmaceutical company promotional activities.

TREATMENT OUTCOMES AND INDICATIONS

A primary influence on prescribing may be the natural desire to achieve the best possible treatment outcome for the patient. For example, if the starting dose of the drug of first choice is not effective in a given patient, the prescriber may choose to increase the dose, add another drug, or switch to a different medication. Sometimes, all of these options will be tried in sequence. For many disorders, the intensity of treatment is titrated against a measured response, such as the blood pressure, blood cholesterol measurement, or the distance that the patient can walk before developing anginal pain. As a result, individuals with more severe underlying disease or more resistant symptoms will tend to receive higher doses of drugs, and greater numbers of drugs. In pharmacoepidemiology studies, it may therefore be difficult to determine whether a given disease–drug association is caused by the drug under study, or is confounded by the nature or severity of the underlying disease state (see also Chapters 16, 20, and 21).

The occurrence of adverse events, for example cough with ACE inhibitors or gastrointestinal bleeding with NSAIDs, will clearly cause prescribers to alter drug choices and to avoid the future use of such agents in the affected patients, and perhaps in other patients. Similarly, the existence of contraindications to certain drugs, like β -blockers in asthma, or penicillin allergy, will impact on prescribers' drug choices for certain patients. Underlying pathology frequently directs drug choice—for example, ACE inhibitors are a reasonable first choice for the treatment of hypertension in diabetic patients, but would be regarded by many as an unnecessarily expensive first drug for newly diagnosed simple hypertension in an otherwise well individual. In the absence of information about diagnosis, other pathology, and contraindications, the accurate interpretation of drug use patterns observed in pharmacoepidemiology studies may be difficult.

EXPECTATION AND DEMAND

Patient demand and expectation have been cited as influencing doctors' decisions to prescribe. However, it appears a gap exists between patients' expectations of a prescription and doctors' perceptions of their expectations. After controlling for the presenting condition, patients in general practice who expected a prescription were up to three times more likely to receive one than those who did not. However, patients whom the general practitioner believed expected a prescription were up to ten times more likely to receive one. It is speculated that failure to ascertain patients' expectations is a major reason why doctors prescribe more drugs than patients expect. Other factors that influenced the decision to prescribe in these studies included the doctor's level of academic qualification, practice prescribing rates, patient exemption from prescribing charges, and difficult consultations.

PERCEPTION OF HARMS AND BENEFITS OF TREATMENT

The harms and benefits of treatment may exert influence on prescribing decisions—patients perceived to be at risk of unwanted adverse effects of therapy are less likely than those without such risks to receive treatment. Perception of harm and benefit may vary with the prescriber. For example, it has been found that compared with cardiologists, general physicians overestimate the benefits of certain cardiac treatments.

Information framing, that is, the manner of presentation of risks and benefits, may influence prescribing decisions. Treatment outcomes presented in terms of relative

risk reduction are more likely to elicit a decision to treat than those presented in terms of absolute risk reduction or as numbers needed to treat. Promotional materials from pharmaceutical companies frequently present the benefits of treatments in relative as opposed to absolute terms, as do the newspaper articles that quote them. As relative effect measures usually appear more striking than absolute measures, relative effect measures may be judged sufficiently impressive to persuade prescription by prescribers too busy to consider the original data in detail. While the decision to prescribe based on such evidence may be justifiable in cases where the absolute benefit happens to be reasonable, inappropriate prescribing decisions may be made if it is very small or insignificant.

Patients may also be influenced by the manner of presenting data on benefits and harms of treatments. For instance, surgery is more likely to be preferred over medical treatment if results are expressed in a positive frame (survival) than a negative frame (mortality).

All human decisions are subject to cognitive biases. As Greenhalgh *et al.* have pointed out, "these biases include anchoring against what is seen to be 'normal', inability to distinguish between small probabilities, and undue influence from events that are easy to recall. Stories (about the harmful effects of medicines) have a particularly powerful impact, especially when presented in the media as unfolding social dramas."

ECONOMIC INFLUENCES

Economic influences, exerted from various sources, may influence drug use and therefore the interpretation of pharmacoepidemiology studies (see also Chapter 22). As medicines, particularly new ones, become increasingly expensive, budgetary restrictions, or indeed incentives, may impact upon prescribing decisions. For example, in 1993, the German government placed a limit on reimbursable drug costs and announced that a proportion of spending in excess of this limit would be recouped from doctors' remuneration budgets. The changes in prescribing patterns, at least in the early aftermath of the limit, were significant. The numbers of prescriptions fell and there was a move to the use of both generic products and older, less expensive drugs.

In England the Department of Health introduced several schemes intended to contain the costs of National Health Service prescribing. These included setting indicative prescribing budgets for general practices, offering incentives to make prescribing savings, and fundholding schemes whereby practices hold and manage their own budgets for a number of services, including prescribing. The effects on

prescribing patterns have been variable. In Australia, pharmaceutical companies are required to provide evidence of the cost-effectiveness of their product, compared with that of an existing alternative, prior to listing on the national list of reimbursable medicines. Generic substitution is encouraged and the costs of "me too" drugs are controlled, in part, by reference pricing. New Zealand also uses pharmacoeconomic analysis and reference pricing, but is unusual among developed countries in also tendering for some of its pharmaceutical needs. This has been highly unpopular among major brand name manufacturing companies.

Other approaches intended to contain prescribing costs have included national formularies and limited lists, patient co-payments, and practice guidelines.

While the approaches outlined above reflect some attempts of governments to contain drug costs by influencing prescribing choices, patients themselves may also exert influence based upon their ability to pay for medicines. Where patients are covered by state or private insurance schemes, medicine expense may not be perceived by the patient or the prescriber to be an issue and drug choice will not be constrained by ability to pay. In fact, more expensive choices than are absolutely necessary may be encouraged. However, for patients required to pay in whole or in part for their medicines, costs may well influence drug choice and, for instance, a diuretic as opposed to an ACE inhibitor or calcium antagonist may be chosen for hypertension treatment, although not necessarily the best choice for the individual concerned. Even in countries with strong social insurance programs patients sometimes have difficulty in affording medications, because of relatively high patient co-payment levels.

Prescribers themselves may have a pecuniary interest in prescribing. Fee for service methods of physician remuneration (as against a capitation fee) have been found to encourage a higher use of services. In Japan, physicians dispense as well as prescribe medicines and the associated financial incentive is considered to contribute to the high numbers of prescriptions per capita and the use of expensive drugs. Concerns about the effects on prescribing of incentives offered to doctors by the pharmaceutical industry have led to such practices being discouraged in most countries and manufacturers have voluntarily adopted a code of good promotional practice.

THE PHARMACEUTICAL INDUSTRY

Promotional activities of the pharmaceutical industry can affect prescribing practices in ways that are relevant to pharmacoepidemiology. For example, if a manufacturer

promotes a new NSAID as being less prone to cause gastrointestinal toxicity than other NSAIDs, it may be given to individuals who have an intrinsically higher risk of gastrointestinal bleeding, such as those who have developed dyspepsia while receiving other NSAIDs, or who have a past history of gastric ulceration. These individuals would therefore be expected to have an increased risk of subsequent gastrointestinal bleeding in comparison with those receiving other NSAIDs, although such a finding might be wrongly attributed to the new drug. This form of “channeling” has been a particularly strong feature with COX-2 inhibitors and, if not adjusted for in nonrandomized studies, will give a misleadingly pessimistic impression of the gastrointestinal toxicity of this class of drugs.

Pharmaceutical companies may exert influence, direct and indirect, on prescribing choices. This may occur through their representatives who visit doctors to provide information about drug products on a one-to-one basis, the sponsorship of educational meetings, the employment of personnel (e.g., nurses at asthma or diabetic clinics), sponsorship to attend international specialty meetings, or invitations to specialists to become “expert advisors” in their particular areas of practice.

PATIENT BEHAVIOR

Consumer behavior must also be considered in pharmacoepidemiology studies. Numerous studies have shown that individuals with some diseases, particularly illnesses that are asymptomatic, such as hypertension and hypercholesterolemia, tend to have poor compliance with prescribed drug therapy regimens. Therefore, if a pharmacoepidemiology study were to be performed in such a situation, and the use of a drug were operationally defined as the dispensing of a prescription, then the number of prescriptions dispensed might overestimate the true exposure to that medication. On the other hand, compliance with some drugs, such as oral contraceptives, tends to be good because consumers are highly motivated to take them. In the case of drugs that are taken for particular symptoms, such as pain or wheezing, individuals may take more medication than is prescribed. If this occurs chronically, it should be reflected in the number of prescriptions that have been dispensed for an individual over a given time period.

The use of non-prescription drugs, which sometimes have the same effects as prescription drugs, also needs to be considered. For example, when examining the effects of NSAIDs using prescription data, it is important to consider the possibility that individuals who appear to be unexposed might actually have been exposed to a non-prescription

NSAID. There is a general trend worldwide for a wider variety of drugs, previously only available on prescription, to become available over-the-counter. In many countries, drugs known to have significant potential for causing interactions, such as cimetidine, are included in this non-prescription availability. Of course, the increasing use of dietary supplements, with virtually no quality control or effective regulation, makes this even worse.

Consumers of prescribed medications may differ from nonusers in a number of other ways that may confound pharmacoepidemiology studies, for example, alcohol intake and smoking status. Unfortunately, this information is rarely, if ever, available from some data sources (e.g., automated databases). Individuals who take certain drugs may use other medical services or have different lifestyles from nonusers. In the case of post-menopausal estrogen therapy, consumers were shown to make greater use of other medical services and to have higher levels of exercise than nonconsumers. This is important, because these factors were potential confounders of the relationship between estrogen use and outcomes such as hip fracture and myocardial infarction (see Chapter 21).

Knowledge of prescriber and consumer behavior is crucial when conducting pharmacoepidemiology studies. Both high doses of drugs and the use of drug combinations are often markers for more severe underlying diseases. Therefore, attempts to link exposure to a drug with a particular outcome must take account of these factors. Disease severity or intolerance to previous medications may be linked in subtle ways to the outcomes of interest, and pharmacoepidemiology studies are subject to these forms of confounding. Economic and promotional influences may affect prescribing patterns in a number of ways, both obvious and subtle, and also require consideration as potential confounders.

CONCLUSIONS

Pharmacoepidemiology is a complex and inexact science. It would be convenient if: exposures and outcomes could always be assumed to be dichotomous; the relationships could be assumed to be unconfounded; and if risk could be assumed to increase proportionately with duration of exposure. However, because of the complexity of the use and effects of drugs among the population, these simplifying assumptions are often violated. Users of drugs will often differ in many respects from nonusers, and in ways that are not easily adjusted for. These differences may confound the associations between exposure and outcomes. Responses to drugs are very variable, not only between individuals but

also within individuals over time. This variability in inter- and intra-individual responses can result in adverse reactions being manifest early in treatment, and the development of tolerance in long-term users. A study of clinical pharmacology provides us with many insights, and a knowledge of the underlying principles is essential during the conduct, and particularly the interpretation, of pharmacoepidemiology studies.

Key Points

- Clinical pharmacology comprises all aspects of the scientific study of medicinal drugs in humans. It can be divided into pharmacokinetics (the relationship between the dose of a drug administered and the serum or blood level achieved) and pharmacodynamics (the study of the relationship between drug level and effect).
- There are many factors that affect an individual's response to a drug. These factors include sex, age, health conditions, concomitant medications, and genetic makeup. An important goal of pharmacoepidemiology is to use population research methods to characterize factors that influence individual drug response.
- Factors that influence individual drug response may do so via pharmacokinetic mechanisms, pharmacodynamic mechanisms, or both.

SUGGESTED FURTHER READINGS

Birkett D. Drug protein binding. *Aust Prescr* 1992; **15**: 56–7.
 Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron* 1976; **16**: 31–41.

Greenhalgh T, Kostopoulou O, Harries C. Making decisions about benefits and harms of medicines. *BMJ* 2004; **329**: 47–50.
 Huang Y-S, Chern H-D, Su W-J, Wu J-C, Chang S-C, Chiang C-H *et al*. Cytochrome P450 2E1 genotype and the susceptibility to antituberculosis drug-induced hepatitis. *Hepatology* 2003; **37**: 924–30.
 Ingelman-Sundberg M. Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. *J Intern Med* 2001; **250**: 186–200.
 Klotz U, Avant GR, Hoyumpa A, Schenker S, Wilkinson GR. The effect of age and liver disease on the disposition and elimination of diazepam in adult man. *J Clin Invest* 1975; **55**: 347–59.
 MacDonald TM, Movant SV, Robinson GC, Shield MJ, McGilchrist MM, Murray FE, McDevitt DG. Association of gastrointestinal toxicity of nonsteroidal anti-inflammatory drugs with continued exposure: a cohort study. *BMJ* 1997; **315**: 1333–7.
 Meyer UA. Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annu Rev Pharmacol Toxicol* 1997; **37**: 269–96.
 Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J *et al*. Coverage by the news media of the benefits and risks of medications. *N Engl J Med* 2000; **342**: 1645–50.
 Ranek L, Dalhoff K, Poulsen HE, Brosen K, Flachs H, Loft S *et al*. Drug metabolism and genetic polymorphism in subjects with previous halothane hepatitis. *Scand J Gastroenterol* 1993; **28**: 677–80.
 Rang HP, Dale MM, Ritter JM, Moore PK. *Pharmacology*, 5th edn. Edinburgh: Churchill Livingstone, 2003.
 Royal Australasian College of Physicians. Available from: <http://www.racp.edu.au/>.
 Shenfield G, Le Couter DG, Rivory LP. Updates in medicine: clinical pharmacology. *Med J Aust* 2002; **176**: 9.
 Wiholm BE, Alvan G, Bertilsson L, Sauve J, Sjoqvist F. Hydroxylation of debrisoquine in patients with lactic acidosis after phenformin. *Lancet* 1981; **I**: 1098–9.

5

When Should One Perform Pharmacoepidemiology Studies?

Edited by:

BRIAN L. STROM

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

As discussed in the previous chapters, pharmacoepidemiology studies apply the techniques of epidemiology to the content area of clinical pharmacology. This chapter will review when pharmacoepidemiology studies should be performed. It will begin with a discussion of the various reasons why one might perform pharmacoepidemiology 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 PHARMACOEPIDEMOLOGY STUDIES

The decision to conduct a pharmacoepidemiology 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 pharmacoepidemiology 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. Regardless, with the exception of postmarketing randomized clinical trials, the cost per patient 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 pharmacoepidemiology 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 pharmacoepidemiology studies could be conceptualized in four different categories: regulatory, marketing, legal, and clinical (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 pharmacoepidemiology study is regulatory: a plan for a postmarketing pharmacoepidemiology study is required before the drug will be approved for marketing. Requirements for postmarketing research have become progressively more frequent in recent years. In fact, since the early 1970s the FDA has required postmarketing research at the time of approval for about one third of all newly approved drugs. Many of these required studies have been randomized

clinical trials, designed to clarify residual questions about a drug's efficacy. Others focused on questions of drug toxicity. Often it is unclear whether the pharmacoepidemiology 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 and prazosin. These are discussed more in Chapters 1 and 2.

Sometimes a manufacturer may offer to perform a pharmacoepidemiology 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

Table 5.1. Reasons to perform pharmacoepidemiology studies

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

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.

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 pharmacoepidemiology 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, following reports about this problem to the FDA's Spontaneous Reporting System.

Finally, drugs are obviously marketed at different times in different countries. A postmarketing pharmacoepidemiology study conducted in a country which marketed a drug relatively early could be useful in demonstrating the safety of the drug to regulatory agencies in countries which have not yet permitted 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, pharmacoepidemiology studies are performed primarily to obtain the answers to clinical questions. However, it is clear that a major underlying reason for some pharmacoepidemiology 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 pharmacoepidemiology 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 pharmacoepidemiology studies that recruit subjects for the study via prescribers. However, as discussed in Chapter 13, 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. 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.

Pharmacoepidemiology studies can also be useful to reposition a drug that is already on the market, i.e., 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 22) and on patients' quality-of-life (see Chapter 23). 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 21), 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 pharmacoepidemiology studies.

Finally, and perhaps most importantly, pharmacoepidemiology 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 pharmacoepidemiology studies in anticipation of this problem. Sometimes these problems can be specifically foreseen and addressed. More commonly, they are not. However, the availability of an existing cohort of exposed patients and a control group will often allow a much more rapid answer than would have been possible if the study had to be conducted *de novo*. One example of this is

provided by the experience of Pfizer Pharmaceuticals, when the question arose about whether piroxicam (Feldene®) was more likely to cause deaths in the elderly from gastrointestinal bleeding than the other nonsteroidal anti-inflammatory drugs. Although Pfizer did not fund studies in anticipation of such a question, it was fortunate that several pharmacoepidemiology research groups had data available on this question because of other studies that they had performed. McNeil was not as fortunate when questions were raised about anaphylactic reactions caused by zomepirac. If the data they eventually were able to have had been available at the time of the crisis, they might not have removed the drug from the market. More recently, Syntex recognized the potential benefit, and the risk, associated with the marketing of parenteral ketorolac, and chose to initiate a postmarketing surveillance cohort study at the time of the drug's launch. 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. One often hears the phrase "What you don't know, won't hurt you." However, in pharmacoepidemiology this view is short-sighted 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 pharmacoepidemiology 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 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, piroxicam, zomepirac, and ketorolac 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. 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 21).

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.

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.

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.

The formulation of the drug can be considered a determinant of the need for formal screening pharmacoepidemiology

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.” 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.

The former is the focus of much of pharmacoepidemiology and most of this book. The latter is the focus of the following discussion.

RISK TOLERANCE

Whether or not to conduct a postmarketing surveillance pharmacoepidemiology 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

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
-

question of the excess risk (see Chapter 2). Use of tampons is extraordinarily strongly linked to toxic shock: the relative risk appears to be 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 toxic shock syndrome in those who use tampons.

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 which 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 a "dread disease."

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 has probably slowed the success of anti-smoking 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 six-year delay in introducing beta-blockers into the US market has been blamed for resulting in more deaths than all recent adverse drug reactions combined. 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. 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.

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.

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. Yet, these deaths are accepted with little concern, while the uncommon risk of an airplane crash or being struck by lightning generates fear. The decision about whether to allow isotretinoin to remain on the market hinged on whether the efficacy of the drug for a small number of people who had a disease which was disfiguring but not life-threatening was worth the birth defects that would result in some other individuals. There is no way to remove this subjective component from the decision about the acceptability of risks. Indeed, much more research is needed to elucidate patients' preferences in these matters. However, this subjective component is part of what makes informed consent so important. Most people feel that the final subjective judgment about whether an individual should assume the risk of ingesting a drug should be made by that individual, after education by their physician. However, as an attempt to assist that judgment, it is useful to have some quantitative information about the risks inherent in some other activities. Some such information is presented in Table 5.3.

Table 5.3. Annual risks of death from some selected hazards^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

^aData derived from Urquhart and Heilmann (1984), O'Brien (1986) and Silverberg and Lubera (1988).

CONCLUSION

This chapter reviewed when pharmacoepidemiology studies should be performed. After beginning with a discussion of the various reasons why one might perform pharmacoepidemiology 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 pharmacoepidemiology study, the next part of this book will provide perspectives on pharmacoepidemiology from some of the different fields that use it.

Key Points

- The decision to conduct a pharmacoepidemiology 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 pharmacoepidemiology study are the costs (monetary, effort, time) of conducting the study itself,

(Continued)

the opportunity costs of other research that might be left undone if this research is performed, 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, and that it could provide false reassurances about a drug's safety.

- The benefits of pharmacoepidemiology studies could be conceptualized in four different categories: regulatory, marketing, legal, and clinical. 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.
- There are several factors that can affect one's willingness to tolerate the risk of adverse effects from drugs. 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.

SUGGESTED FURTHER READINGS

- Binns TB. Therapeutic risks in perspective. *Lancet* 1987; **2**: 208–9.
- Bortnichak EA, Sachs RM. Piroxicam in recent epidemiologic studies. *Am J Med* 1986; **81**: 44–8.
- 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.
- 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.
- Humphries TJ, Myerson RM, Gifford LM *et al*. A unique post-market outpatient surveillance program of cimetidine: report on phase II and final summary. *Am J Gastroenterol* 1984; **79**: 593–6.
- Joint Commission on Prescription Drug Use. *Final Report*. Washington, DC, 1980.
- Lowrance WW. *Of Acceptable Risk*. Los Altos, CA: William Kaufmann, 1976.
- Marwick C. FDA ponders approaches to curbing adverse effects of drug used against cystic acne. *JAMA* 1988; **259**: 3225.
- Mattison N, Richard BW. Postapproval research requested by the FDA at the time of NCE approval, 1970–1984. *Drug Inf J* 1987; **21**: 309–29.
- O'Brien B. "What Are My Chances Doctor?"—A Review of Clinical Risks. London: Office of Health Economics, 1986.
- Rogers AS, Porta M, Tilson HH. Guidelines for decision making in postmarketing surveillance of drugs. *J Clin Res Pharmacol* 1990; **4**: 241–51.
- Rossi AC, Knapp DE. Tolmetin-induced anaphylactoid reactions. *N Engl J Med* 1982; **307**: 499–500.
- Silverberg E, Lubera JA. Cancer statistics. *CA Cancer J Clin* 1988; **38**: 5–22.
- Stallones RA. A review of the epidemiologic studies of toxic shock syndrome. *Ann Intern Med* 1982; **96**: 917–20.
- 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.
- 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.
- 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.
- 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.
- Urquhart J, Heilmann K. *Risk Watch—The Odds of Life*. New York: Facts on File, 1984.
- Young FE. The role of the FDA in the effort against AIDS. *Public Health Rep* 1988; **103**: 242–5.

6

Views from Academia, Industry, and Regulatory Agencies

The following individuals contributed to editing sections of this chapter:

LEANNE K. MADRE,¹ ROBERT M. CALIFF,² ROBERT F. REYNOLDS,³ PETER ARLETT,⁴
and JANE MOSELEY⁵

¹Duke University Medical Center, Durham, North Carolina, USA; ²Duke Clinical Research Institute, Durham, North Carolina, USA; ³Pfizer Inc., New York, NY, USA; ⁴Pharmaceuticals Unit, DG Enterprise and Industry, European Commission, Brussels, Belgium; ⁵Medicines and Healthcare Products Regulatory Agency, London, UK.

This chapter presents different perspectives on pharmacoepidemiology, in particular the perspectives of academia, industry, and regulatory agencies.

A VIEW FROM ACADEMIA

The field of pharmacoepidemiology provides a challenge to the traditional academic community. It may be viewed as one element of a group of disciplines necessary to understand how to deliver diagnostic and therapeutic technologies in a manner that optimizes health—a field that might more broadly be called “therapeutics.” Various forces continue to push issues of the risks and benefits of therapy into the public consciousness, while academic medicine has had difficulty accepting that the discipline should be a focus. This reticence to embrace the study of therapeutics as a priority relative to the more basic sciences is one element

of a stance taken by academic medical centers (AMCs) that has contributed to a backlash about the size of the public investment requested to support the research done in academia.

However, academia is fully capable of creating new approaches that can provide a basis for the discipline of therapeutics. The Centers for Education and Research on Therapeutics (CERTs) organization represents a US example of an effort to change this dynamic by creating a consortium of academic centers linked to multiple government and private entities with a vision of serving as a trusted national resource for people seeking to improve health through the best use of medical therapies. This program, mandated by the authorization for the FDA, brings together AMCs, government agencies, the medical products industry, and consumer advocates with core funding through the Agency for Healthcare Research and Quality (AHRQ) and offers the opportunity to join government with private funding to

meet the mission. The mission of the CERTs is to conduct research and provide education that will advance the optimal use of drugs, medical devices, and biological products. This mission is achieved through activities that develop knowledge about therapies and how best to use them, manage risk by improving the ability to measure both beneficial and harmful effects of therapies as used in practice, improve practice by advancing strategies to ensure that therapies are used always and only when they should be, and inform policy makers about the state of clinical science and effects of current and proposed policies. This multicenter effort is at least partially succeeding in bringing pharmacoepidemiology, and the study of therapeutics in general, back into the mainstream of academic medicine, and it provides insight into additional approaches that are needed.

ISSUES DRIVEN BY SUCCESS

People are living longer with less disability than ever before. Whereas most previous gains came from broad public health measures, an increasing portion of the gains in disability-free life expectancy comes from medical care. While we are making steady progress in therapeutics, it can be accelerated by better planning and integration of the discipline of clinical therapeutics, and AMCs can contribute greatly to these efforts.

The basis for a growing gap between the potential and the reality of therapeutics is the confluence of multiple societal trends. The United States and other postindustrial countries are experiencing a dramatic change in demographics, with an enormous increase in the proportion of the population that will be elderly. At the same time, in developing countries, progress is being made in instituting public health, economic, and educational measures that will reduce the epidemic causes of premature death. These changes will greatly increase the importance of medical therapeutics to prevent, delay, treat, and palliate chronic diseases. The two new issues on the scene, massive obesity in the young and biological terrorism, only heighten the importance of therapeutic knowledge and academic infrastructure to supply a competent and creative workforce.

These demographic and public health changes are occurring at a time when a revolution in biological knowledge is leading to previously unthinkable therapeutic possibilities. The combined research investment of the US government and the medical products industry (drugs, biologics, and devices) now exceeds \$60 billion per year. Most common diseases have one or more known effective therapies, and many highly prevalent diseases such as cardiovascular disease and cancer have multiple effective therapies.

The continued evolution of computing has enabled pharmacoepidemiology in particular to take an even more prominent role. Publications or news stories about secondary analyses of data sets showing a relationship between a therapy and an outcome abound because health care transactions are increasingly captured in computers, either directly or by the capture of claims data, and these data sets can be manipulated by increasingly usable and powerful statistical packages. This technological advance has also rapidly opened up cross-cultural communications, thus fostering international collaborations exemplified by the International Society for Pharmacoepidemiology (ISPE) (www.pharmacoepi.org).

All of these trends are positive, but they raise a new set of problems that must be addressed, at least partly through the efforts of AMCs. As the potential of technology continues to accelerate and our inability to provide all technologies to all people is increasingly evident, we need better knowledge about how to effectively apply technologies, including diagnostic devices, drugs, biologics, and therapeutic devices, to the right patient at the right time.

Inadequate Knowledge Base

There is a large and growing gap between the potential of therapies to ameliorate human disease and our actual base of knowledge. This gap does not emanate from lack of progression in studies of therapeutics. Rather, the problem has grown because the pace of technology development continues to accelerate relative to our ability to test and evaluate it.

Drugs and devices are still developed in relatively small studies of limited duration, often without measuring health outcomes as an endpoint and almost always without direct measures of costs. Instead we rely on biomarkers, surrogates, or partial efficacy evaluations for regulatory approval, in combination with limited safety data from short-term studies. These studies usually also exclude many patients who ultimately will receive these therapies. The excluded patients tend to be the elderly and minorities, usually with a high rate of comorbidities, particularly renal dysfunction or advanced chronic disease. The result is that many therapeutics reach the market with incomplete documentation of their use in the populations at highest risk.

Suboptimal Practice

The gap between the practice of health care delivery and the knowledge base that should guide that practice is vast. In primary prevention and in diseases for which standards of care have evolved, demonstrations of imperfect levels of adherence to practice standards abound. This gap has been well demonstrated in the hospital setting, where decision

making is largely up to the treating physician. Evolving data from the outpatient setting show an even larger gap. The complexity of the outpatient therapeutic transaction is driven by multiple factors, most prominently the actions of the patient in adhering to the prescriber's recommendations.

The CERT at the University of Alabama at Birmingham has focused on inadequate practices in the area of osteoporosis treatment, leading to substantial collaborative efforts to develop guidelines and quality indicators in rheumatology. In addition, a study of antibiotic prescribing at the CERT at the University of Pennsylvania has shown the tension between the desire of practitioners to improve societal outcomes and their obligation to individual patients. By combining the analytic and quantitative skills in academic centers and the broad desire of practitioners to improve, particularly in a managed care setting, we are seeing an improvement in practice nationally.

Counterproductive or Nonproductive Policies

At the time of the initial funding of the CERTs, numerous counterproductive policies relating to therapeutics could be cited. The prototypical case was the temporary effort in the state of New Hampshire to limit the use of prescription medications in the state Medicaid program. The result was that the emergency departments in the state were overwhelmed with acute visits from people whose chronic diseases had become uncontrolled due to lack of access to drugs.

The CERT at the University of Pennsylvania, using sophisticated epidemiological techniques, found no benefit from the usual state Medicaid approach with regard to prescribing errors or avoidable hospitalizations (see also Chapter 27). The investigators believed that the futility of applied programs such as low rates of alerts, lack of linkage between alerts, the complex reasons that drugs are prescribed, and the time lags between prescription and review can probably be explained by structural and functional principles. The ability of several academic centers to collaborate on assessing the evidence is likely to lead to further improvement in advice to policy makers. (See Case Example 6.1.)

CASE EXAMPLE 6.1: THE ROLE OF ACADEMIC MEDICAL CENTERS IN MEDICAL THERAPEUTICS

Background

- There is a knowledge deficit related to the benefits and risks of marketed therapeutics.

- Documented gaps exist between what is known and what is practiced.
- Health care transactions are increasingly captured in computers, enabling pharmacoepidemiology to have a more prominent role in understanding therapies.

Question

- What roles do Academic Medical Centers (AMCs) have in ensuring therapeutics are used to maximize patient benefit and minimize the likelihood of harm?

Approach

- AMCs can create repositories of population database resources for the purpose of developing a generalized understanding of diagnostic and therapeutic strategies.
- AMCs train future practitioners and researchers; develop and evaluate therapies; deliver care to patients; and have thought leaders who influence national policy.
- AMCs can leverage their unique position to move national practice towards higher quality health care.
- The Centers for Education and Research on Therapeutics (CERTs) are a consortium of AMCs linked to multiple government and private entities with a vision of serving as a trusted national resource for people seeking to improve health through the best use of medical therapies.

Results

- Better and more information about benefits, risks, and appropriate use of therapeutics.
- Increased likelihood that health care providers understand and practice evidence-based care.

Strengths

- Pharmacoepidemiology is a necessary discipline to understand how to deliver diagnostic and therapeutic technologies in a manner that optimizes health.
- AMCs are uniquely qualified to create an infrastructure that bridges the gap between therapeutics knowledge and practice.

Limitations

- Requires cooperation among different expertise and perspectives within a fragmented health care system.
- Knowledge alone is insufficient to ensure therapeutics are used in such a way as to maximize benefit and minimize harm to patients.

(Continued)

Summary Points

- New issues, such as biological terrorism, heighten the importance of therapeutic knowledge and academic infrastructure to supply a competent and creative workforce.
- The CERTs organization represents an approach to creating a therapeutics discipline.
- Pharmacoepidemiology plays an increasingly important role relative to understanding therapeutics benefits, risks, and appropriate use.

ROLE OF ACADEMIC MEDICAL CENTERS

While AMCs may be only one of many entities in the therapeutics arena, all physicians, pharmacists, and dentists are trained at AMCs, as are a large proportion of nurses. Therefore, AMCs have the opportunity and responsibility for providing the national infrastructure for practitioners of therapeutics. AMCs also provide a home for the majority of clinicians who influence public sources of information and opinion, and receive the bulk of funding from the US National Institutes of Health (NIH), the main source of medical research funding.

Improve Knowledge

While knowledge alone is not enough to correct the gap between research and practice, it is a necessary first step. This effort begins with training in nursing, pharmacy, medical, dental, and public health schools.

Much of the work of the CERTs has been focused on improving the knowledge of therapeutics among practitioners, but AMCs have not done an adequate job of relating to practitioners, leaving much knowledge transmission to the medical products industry itself. New legislation in the United States has given the opportunity to take back a large portion of continuing education into the AMC arena. Professional organizations, buttressed by government regulation about using continuing medical education for advertising, have produced stringent guidelines that call for independent control of programs that educate health professionals.

Reinvigorate Clinical Pharmacology

The field of clinical pharmacology has been poorly supported by federal funding sources, and for the most part AMCs have responded by decreasing the number of faculty positions. This mismatch of funding and societal need can be

corrected partially by programs such as the CERTs. In addition, as the importance and scope of clinical pharmacology have expanded, several key entrepreneurial opportunities have emerged. The General Clinical Research Centers funded by the NIH have continued to receive excellent levels of funding. The pharmaceutical, biotechnology, and device industries are also feeling the pressure from the shortage of talent in developing the knowledge needed for successful drug development. The focus on the safety of medical products and the long-term balance of risk and benefit has created a large demand for expertise in the industry, well beyond the focus on product development. Finally, the societal focus on patient safety will require methodology developed by this field to monitor the use of drugs and devices. Further efforts are needed to link AMCs with the needs of the industry and government with regard to the postmarketing phase.

Create a Repository of Database Resources

The evidence needed to guide choices at the individual and policy levels will increasingly be driven by empirical analysis of data from populations. At a national level, entities such as the CERTs can provide the opportunity for multiple parties to answer questions from databases. Multiple databases are available for interrogation with specific questions about therapeutics (see Chapters 11 and 12). However, recent concerns about privacy (see Chapter 19) and the sheer size of the databases emphasize the need for facilities with experts in data management and analysis. The CERTs represent one approach to this by bringing together experts from multiple academic centers into an organization with a coordinating center charged with facilitating common projects.

In this construct the databases continue to reside with their developers, but systems are developed to enhance the sharing of portions of the databases to answer specific questions. The HMO Research Network (see Chapter 12) has created a prototypical example of this “federated” approach, in which questions can be asked, and the system brings together elements of databases from the various health maintenance organizations to provide answers. Issues of privacy, anonymization, sharing, and scope of the data have been worked out in advance to improve the efficiency of the process of gaining access to the data.

At the local level, there is a pressing need for institutions and health systems to develop data repositories and local expertise in appropriate analysis related to quality. AMCs have the expertise to set the example by organizing data repositories and providing access to data for the purposes of improving quality and developing a generalized

understanding of diagnostic and therapeutic strategies. This effort should not stop at the national level, however. Indeed, access to the UK General Practice Research Database (see Chapter 12) has allowed multiple important observations to be made of direct relevance to global health.

Improve Practice

While AMCs train providers in their basic skills, the province of clinical practice has many entities with variable agendas and opportunities to demonstrate better approaches to the organization of health care. This larger clinical enterprise dwarfs the ability of AMCs to directly change the practice of medicine. In this regard, AMCs should seek to leverage their unique position to move national practice towards higher quality health care.

At the most fundamental level, AMCs have a responsibility to apply resources to science by evaluating therapies that are already marketed. The medical products industry lacks the incentive to study its products as patents' lives near their end. Furthermore, rules on generics require only evidence of bioequivalence, and the profit margin on generics is not thought to justify the funding of outcome studies. A similar situation exists with regard to food additives and "alternative and complementary" therapies. There is no source to perform these evaluations other than the cadre of AMCs, ideally buttressed with federal funding. The Clinical Research Roundtable of the Institute of Medicine has stressed the need for an elevation of funding for pragmatic clinical trials.

AMCs also should be instilling in practitioners a fundamental understanding of the principles of therapeutics and the measurement of quality in health care delivery. Although the American Society for Clinical Pharmacology and Therapeutics has put forward a model curriculum for its particular domain, a well-defined curriculum on the broader field of therapeutics does not exist, and teaching of the fundamentals of biostatistics, probability, decision making, and health care systems has not been well received, and perhaps not well executed. The US curriculum is packed, and many contingents are so adamant about not losing time that inserting new material is seen as a zero-sum game, in which adding anything new means eliminating something else. Nevertheless, we are seeing a gradual increase in the emphasis on skills that will improve the quality of therapeutics in undergraduate, house staff, and continuing education programs. One approach to this dilemma has been the development of focused curricula which are posted on the Internet for use in multiple settings and institutions.

AMCs also have the challenge of training and supporting the researchers who will define the field in the future.

Currently, the AHRQ has scant funds for training and faculty development, and this arena has purposefully not been a focus of the NIH. The NIH Roadmap may provide the opportunity to develop a larger cadre of experts in the related fields of clinical epidemiology, biostatistics, clinical trials, health economics, health services research, and clinical pharmacology.

Finally, AMCs must consider developing novel models of care delivery that improve the use of marketed products. Much of the work of the CERTs has indicated that the quality of medical care will improve most dramatically when systematic changes are made in the organization and funding of delivery. The huge increase in people with chronic disease coupled with advances in effective but expensive technology present a challenge for AMCs to lead the way by testing novel approaches to team-based care delivery using advanced information technology.

Inform Constituents

Policy Makers

Ultimately, many issues in therapeutics can only be solved by urging policy makers to increase the chances of rational policies that provide incentives for behavior that improves the use of therapeutics. The potential to leverage expertise for benefit has never been greater as more health care resources fall to the federal government and large payer plans.

The Public

Informing the public about the balance of risks and benefits of therapeutics is a tremendous task, about which surprisingly little is known. An increasing number of studies show that much information made available to the public is either biased or unintelligible. Yet, at AMCs, little attention has been focused on translating medical research findings into statements that can be acted upon effectively by the public.

The Press

Surveys of the public indicate that more health-related decisions are based on press reports than on doctor visits. Even when the FDA wants to get a message to the public, it must use the press to get this message out. It can be argued that observational studies about the broad issue of therapeutics dominate the press reports on medicine, often with inadequate attention to the quantitative issues involved in the interpretation. Unfortunately, the role of the press in therapeutics and public health has received inadequate attention in academia.

SPECIAL ROLE OF PUBLIC–PRIVATE PARTNERSHIPS

The dramatic effects of the aging of the population combined with the fattening of the younger generation will create an enormous societal challenge. The problems are arguably so overwhelming that public–private partnerships may be the only way to create enough resources to find effective solutions.

The CERTs have developed a model approach to public–private partnership based on a set of principles that increase the likelihood that modest public investment will yield a larger private investment while also safeguarding the ecumenical nature of the enterprise. These principles are designed to encourage engagement of industry partners in the research enterprise under a set of rules that can be considered by the CERTs' governance and openly discussed.

The first priority of the CERTs is to tackle issues of public interest and improve the rational use of therapeutics through research and education activities that would not otherwise be done. Second, the CERTs are actively seeking public–private partnership, rather than avoiding it. The CERTs organization is a public–private partnership, therefore centers seek useful, appropriate interactions with private organizations to support and enhance education, research, and demonstration projects. The AHRQ works with the centers to establish appropriate agreements to optimize use and sharing of resources. Third, the issue of conflicts of interest, likely in any public–private partnership, is acknowledged and confronted. The obligation is to disclose fully and manage potential conflicts in a manner that minimizes the risk of those conflicts, while maximizing progress to achieve the CERTs goals. Fourth, academic integrity is paramount. As academic researchers, individuals conducting projects under the CERTs umbrella will retain final decision making about study design, analysis, conclusions, and publication, and will ensure that their work complies with their respective institutions' conflict-of-interest rules.

CONCLUSIONS

The role of epidemiology in drug development, safety assessment, and commercialization has expanded significantly in recent years. Traditionally, the pharmaceutical industry and regulatory agencies relied on basic science research and clinical studies of experimental design to assess the efficacy and safety of new medications before approval, and spontaneous reporting systems to assess the safety of

medications after approval. Fifteen years ago, epidemiology was primarily used defensively in response to legal or regulatory questions. Now, however, pharmaceutical and biotechnology companies employ epidemiologists and apply observational study designs and methods in many functional areas within pharmaceutical companies.

In summary, AMCs have a vital role to play in pharmacoepidemiology and therapeutics. This role includes providing a fundamental basis of training and maintenance of an academic discipline. However, it also includes creative integration with health care providers, government agencies, and the broader medical industry.

A VIEW FROM INDUSTRY

During drug development and marketing, national regulatory agencies, such as the FDA, or multinational agencies, such as the European Medicines Evaluation Agency (EMA), regulate the pharmaceutical industry. These agencies require that a pharmaceutical manufacturer demonstrate that a new medication, device, or vaccine is safe and effective before being approved, and that information about the effects of these medications is communicated to patients and physicians. The manufacturer has a further obligation to evaluate the safety of products on an ongoing basis to develop and maintain product labeling that ensures appropriate prescribing of drugs by physicians and safe use by patients. Implicit in this process is the need for a logical basis for drug approval, a rational and balanced approach to both pre- and post-marketing surveillance of drug safety, and a scientific, evidence-based regulatory environment. As such, manufacturers devote significant efforts and resources to meeting worldwide regulatory requirements for drug research and development, monitoring the postmarketing safety of medications in compliance with required spontaneous reporting systems and time frames, and in completing Phase IV commitments.

Over the past two decades, there has been an increased regulatory emphasis on postmarketing surveillance, and a greater likelihood for observational studies to be required as post-approval safety commitments. Regulatory agencies have long recognized the importance of continuously evaluating the risk/benefit balance of medications. Recently, however, regulatory agencies and the pharmaceutical industry have placed greater importance on the development of guidelines and standard processes for recognizing and, where possible, minimizing risk. There has been a gradual shift from the traditional mode of passive risk assessment and communication (e.g., voluntary spontaneous reporting

systems and information dissemination) to the use of more active forms of evaluating and managing potential medication risks, such as restrictions in use and distribution or mandatory education programs. This shift has resulted in calls for a scientifically-based process for managing medication risks.

Risk management—the scientific process by which risks are identified, assessed, communicated, and minimized—has a formal role in the development, review, and approval of new drugs in the US and Europe, and other regulatory bodies are taking similar steps (see also Chapter 27). New legislation acknowledges that there are both risks and benefits inherent in therapeutic interventions, and that a common goal of manufacturers and regulatory agencies is optimizing therapeutic benefit and minimizing medication risk. Revenue collected in the form of prescription drug user fees will now be earmarked for certain postmarketing risk assessment activities. Further, recommendations that risk management plans be developed prior to drug approval have played a key role in moving the risk management planning process into earlier stages in drug development, even though filing of a formal pre-approval risk management plan remains voluntary at this time. The focus on risk management during the pre-approval period provides an opportunity to explore and quantify potential safety signals and document the exploratory and decision-making process and rationale in a risk management plan, which will necessarily evolve throughout the life cycle of the drug. Epidemiology plays a central role in risk management activities, through studies of the natural history of disease, disease progression/treatment pathways, and mortality and morbidity patterns, the design and implementation of post-approval safety studies, and risk minimization programs. The emerging risk management framework, with its emphasis on scientifically-based methodologies and transparent decision making, provides a unique opportunity for epidemiologists to contribute to the development of effective and safe medications and build the public's confidence in the actions of industry and government.

EPIDEMIOLOGY IN THE PHARMACEUTICAL INDUSTRY

Epidemiology contributes to several important functions within a pharmaceutical company, including product planning, portfolio development, and the commercialization of drugs, but its greatest contribution is in drug safety evaluation.

Evaluating Drug Safety

There are many relevant safety issues that can only be studied through observational epidemiology. Only epidemiologic methods are practical for estimating the incidence of and risk factors for rarely occurring events in large populations exposed to a drug, to study events with a long latency period, or to study cross-generational effects of a drug. While observational epidemiology offers numerous advantages, though, epidemiologic studies should never be viewed in isolation from other data sources when addressing questions of a drug's safety. Results from clinical trials, spontaneous reports, epidemiologic studies, and, where relevant, preclinical data sets should all be evaluated for their potential to address the particular safety question raised, with close consideration given to the unique strengths and limitations of the study designs and data collection methods used.

Clinical Trials

The randomized controlled clinical trial is considered the gold standard methodology to study the safety and efficacy of a drug. However, trials are limited by the relatively small numbers of patients studied and the short time period over which patients are observed. The numbers of patients included in premarketing clinical trials are usually adequate to identify only the most common and acutely occurring adverse events. Typically, these trials have a total patient sample size up to several thousand. 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. Thus, clinical trials are usually only large enough to detect events that occur relatively frequently, and are not intended to address all potential safety issues related to a particular drug. (See also Chapter 3.)

An additional limitation of clinical trials with respect to drug safety is the strict inclusion/exclusion criteria common in these studies. Patients included in pre-approval clinical studies may be the healthiest segment of that patient population. Special groups such as the elderly, pregnant women, or children are frequently excluded from trials. Patients in clinical trials also tend to be treated for well-defined indications, have limited and well-monitored concomitant drug use, and are closely followed for early signs and symptoms of adverse events which may be reversed with proper treatment. In contrast, once a drug is marketed, it is used in a “real-world” clinical context. Patients using the drug

may have multiple comorbidities for which they are being treated simultaneously. Patients may also be taking over-the-counter medications, “natural” remedies, or illicit drugs unbeknown to the prescribing physician. The interactions of various drugs and treatments may result in a particular drug having a different safety profile in a postmarketing setting compared to the controlled premarketing environment. Adherence to medications also often differs between closely monitored trials and general post-approval use.

Spontaneous Reporting Systems

Spontaneous reporting systems are valuable for identifying relatively rare events and providing signals about potentially serious safety problems, especially with respect to new drugs (see Chapters 7 and 8). While there is currently no uniform definition, a signal is generally understood to be a higher than expected relative frequency of a drug-event pair. Depending on the circumstances and the information available on background rates of events in the populations using the drug, definitions of “higher than expected” will vary by drug class, indication, and over time. Ultimately, signals are used to generate hypotheses, which then may be studied through observational or interventional studies as appropriate; however, spontaneous reports must be interpreted within the context of the strengths and limitations of the particular reporting systems. These voluntary reports are subject to many biases and external influences on reporting rates, which are unmeasured and in many cases unmeasurable. Events may be underreported and the decision as to which events are reported can be strongly affected by bias. The effects of these biases differ among drugs and over time. The number of spontaneous reports received most often relates to the length of time a drug has been on the market, the initial rate of sale of the drug, secular trends in spontaneous reporting, and the amount of time a manufacturer’s sales representatives spend with physicians “detailing” the product. Certain types of events seem more likely to be reported, such as those which are serious and/or unlabeled, those that occur rarely in the general population, and those associated with publicity in the lay or professional media. The frequency of reporting varies by drug class and drug company. The number of reports does not equal the number of patients, since events may be reported several times. Most importantly, and a frequently misunderstood point, valid incidence rates cannot be generated from spontaneous reporting systems, since neither the true numerator (the number of adverse events) nor the true denominator (the number of exposed individuals whose events, if they suffered them, would have been reported) is known, and thus relative safety cannot be assessed with any validity.

Additionally, the events reported have an underlying background rate in the population, even in the absence of drug treatment, which may not be known. Media coverage, in particular, has a significant effect on the timing and volume of adverse events voluntarily reported to spontaneous reporting systems. This effect has been documented for various drug classes and adverse events. However, notwithstanding these important limitations, spontaneous reporting systems have been successfully used in numerous circumstances to alert regulatory agencies and manufacturers to a potentially high frequency of serious adverse events in a newly launched drug.

Signal Detection

Signal detection is a rapidly growing field primarily using data collected in voluntary spontaneous reporting systems to enhance the qualitative screening capabilities of expert medical reviewers at pharmaceutical companies and regulatory agencies (see Chapter 8). Typically, medical reviewers have relied on convincing clinical criteria and frequency of events to identify potential signals. Statistical methods have traditionally been underused in analyses of spontaneous reporting data, largely due to the variable quality of the reports and data collection methods; however, these methods are now being employed in an attempt to identify safety signals earlier than has been possible in the past, due to the large, and ever-increasing, volume of spontaneously-reported post-approval safety data.

Three automated signal detection methods—the proportional reporting ratio (PRR), the Bayesian Confidence Propagation Neural Network (BCPNN) method, and the Multi-item Gamma Poisson Shrinker (MGPS) method—are emerging as the most commonly used by regulatory agencies, drug monitoring centers, and pharmaceutical manufacturers, although validation and the practical utility of these methods and the impact of the adverse event coding dictionary used (e.g., MedDRA, WHO-ART, or COSTART) are still being tested.

All three methods are limited by the fact that most of the signals identified using these methods are known associations or represent one or more forms of confounding (e.g., confounding by indication) that bias voluntary reporting systems. These limits may be addressed in the future by using prospective data collection systems for signal detection, which permit population-based estimates of medication-related adverse events and patterns of drug use. These prospective data sets have relatively complete data on individuals affected by adverse events as well as information about the population using the medication, in contrast to spontaneous reporting databases. The use of

population-based data sources to identify drug-event pairs may result in an ability to detect signals more rapidly than would be possible using passive surveillance systems alone, and to apply methods developed for disease surveillance in other databases, such as the tree-based scan statistic. Since the use of structured data sources for signal detection is new, significant research and validation is required in the coming years to determine how population-based surveillance systems will best contribute to pharmacovigilance.

Descriptive Epidemiology Studies

There is increasing recognition within the pharmaceutical industry that a strong epidemiology program in support of drug development is often important for the successful risk management of new medications. Epidemiologic studies conducted prior to product approval are useful for: establishing the prevalence/incidence of risk factors and comorbidity among patients expected to use the new medication; identifying patterns of health care utilization and prescribing of currently approved treatments; and quantifying background rates of mortality and serious nonfatal events.

With the wide availability of computerized health databases (see Chapters 11 and 12), it is now possible to conduct studies across diverse patient populations (e.g., private/public assistance insurance or varying geographical areas) and compare disease rates, examining the effect of differences in clinical practice or access to health care. When these data are available prior to approval, background rates of mortality and morbidity provide an important context for interpreting rare events observed in Phase III clinical trials and spontaneous reports. These data also provide the “real world” estimates necessary to design feasible postmarketing studies. Descriptive epidemiologic studies can also be conducted post-approval to describe new drug users’ characteristics and patterns of use, and may also provide measurements of the drug’s effectiveness at the population level.

In addition to the role that descriptive epidemiology studies play in characterizing the background rates of morbidity and mortality, 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.

Post-approval Safety Studies

During the premarketing phases of drug development, randomized clinical trials involve highly selected subjects and in the aggregate include at most a few thousand patients. These studies are sufficiently large to provide evidence of a beneficial clinical effect and to exclude large increases in risk of common adverse events. However, premarketing trials are rarely large enough to detect small differences in the risk of common adverse events or to reliably estimate the risk of rare events. Identification and quantification of potentially infrequent but serious risks requires larger studies designed to distinguish between the role of background risk factors and the effects of a particular drug on the rate of outcomes. Because of the complexity of design and cost, large controlled trials have not been widely used for the post-marketing evaluation of drugs. Recently, regulators and the medical community have communicated a desire for safety data from the populations that will actually use the drugs in “real-world” clinical practice settings. 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.

Epidemiologic studies can also be 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 anti-ulcer drugs, cimetidine was associated with the highest risk of developing symptomatic acute liver disease.

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 but not the use of the medication itself (see also Chapters 16 and 21). The result, in observational studies, is a form of selection bias, where patients taking a particular medication are selected in a fashion that makes them at unequal risk of the outcome under study. As with any other form of confounding, one can theoretically control for its effects if one can reliably measure the severity of the underlying illness. In practice, though, this is not easily or completely done, especially when a drug may have specific properties affecting the type of patient it is used for within its indication. In these cases, studies using randomization, whether experimental or observational in design, may be necessary. It is in this context that a Large Simple Trial (LST) design may be the most appropriate study design for postmarketing safety evaluation (see Chapter 20). Randomization of treatment assignment is a key feature of an LST, which controls for confounding of outcome by known and

unknown factors. Further, the large study size provides the power needed to evaluate small risks, both absolute and relative. 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 study approximates real life practice. (See Case Example 6.2.)

**CASE EXAMPLE 6.2:
AN INNOVATIVE STUDY DESIGN TO
ASSESS A MEDICINE'S
POST-APPROVAL SAFETY**

Background

- Geodon (ziprasidone), an atypical antipsychotic, was approved for the treatment of schizophrenia by the US FDA in 2001.
- A comparative clinical study of six antipsychotics conducted by the sponsor demonstrated that ziprasidone's QTc 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.
- No serious cardiac events related to QTc prolongation were observed in the NDA clinical database.
- Patients with schizophrenia have higher background rates of mortality and cardiovascular outcomes, regardless of treatment type.
- It is unknown whether modest QTc prolongation results in an increased risk of serious cardiac events.
- The sponsor proposed and designed an innovative study to assess the post-approval cardiovascular safety of Geodon in clinical practice settings.

Question

- Is modest QTc prolongation associated with an increased risk of death and serious cardiovascular events as medicines are used in the "real world"?

Approach

- Randomized, large simple trial to compare the cardiovascular safety of ziprasidone to olanzapine (ZODIAC, or Ziprasidone Observational Study of Cardiac Outcomes).

Results

- Recruiting 18 000 patients from 18 countries in Asia, Europe, Latin America, and North America.
- Broad entry criteria based on approved labeling.
- Random assignment to ziprasidone or olanzapine.
- No additional study monitoring or tests required after randomization.
- Patients followed up during usual care over 12 months.
- As of November 2005, nearly 17 000 patients enrolled and over 10 000 completed.

Strengths

- Random allocation eliminates confounding by indication and other biases.
- Large sample size allows for evaluation of small risks.
- Study criteria reflect "real world" normal practice, maximizing generalizability.

Limitations

- Minimal data collection limits ability to address multiple research questions.
- Subjective endpoints may be difficult to study using this design.
- Large simple trials are resource- and time-intensive.

Summary Points

- Large simple trials, when conducted as randomized, prospective epidemiology studies, are appropriate for evaluating potentially small to moderate risks.
- Large simple trials permit the study of drug safety in a real-world clinical practice setting while controlling for confounding by indication (see Chapter 20 for further discussion).

Long Latency Outcomes

Epidemiologic methods provide the only practical way to study the association between drugs and effects with very long latency periods. Early recipients of human growth hormone (derived from human cadaveric pituitary tissue) were found to have elevated risks of Creutzfeldt-Jacob disease (CJD). Recombinant growth hormone became available in the mid-1980s, but due to the long latent period for CJD, cases continued to be diagnosed well after that time. Epidemiologic studies have documented the association between iatrogenic leukemia and treatment with alkylating agents or epipodophyllotoxins for previous cancers.

Second malignant neoplasms (of the solid tumor type) have also been associated with the use of alkylating agents and anti-tumor antibiotics. Chemotherapy given to children prior to or during the adolescent growth spurt has been associated with slowing of skeletal growth and loss of potential height. Decreased bone mineral density has also been documented following chemotherapeutic treatment in childhood.

Evaluating a Drug's Effects on 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, due to potential risks to the developing fetus and newborn. In addition, most clinical trials that enroll women cease study of pregnant women upon detection of pregnancy. Thus, at the time of introduction to market, the effects of many medications on pregnancy are not well known, with the foundation of drug safety during pregnancy often resting largely on animal reproductive toxicology studies. This is a significant public health consideration, particularly if the medication will be used by many women of childbearing potential, since approximately half of all pregnancies in the US are unplanned. 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.

Epidemiologic methods have also 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 27). Although animal teratology testing is part of the pre-approval 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.

In certain circumstances registries are used to obtain information about the safety of new medications during pregnancy. 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.

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 (see Chapters 15, 16, and 27). 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 be population-based, thus increasing generalizability. It will allow for a robust assessment between drug exposure and outcome by being prospective in nature, by collecting information on the timing of drug exposure, detailed treatment schedule, and dosing, by using standard and predefined definitions for pregnancy outcomes and malformations, and by recording these data in a systematic manner. The registry will ideally also follow offspring of medication-exposed women for a prolonged period after birth, to allow for detection of any delayed malformations in children who seem normal at birth. 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, i.e., prior to knowledge of pregnancy outcomes, should be separated from those cases identified retrospectively, i.e., after pregnancy outcome has been determined by prenatal diagnosis, abortion, or birth, as the latter may be biased towards reporting of abnormalities (see Chapter 27). To minimize ascertainment bias, risk 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 followed aggressively to obtain complete pregnancy outcome reports.

Portfolio Planning and Development

Product planning is a critical function within innovative pharmaceutical companies, because of the need for new

and promising developmental drugs in product pipelines. Epidemiology plays a key role in the planning and development process. For example, basic epidemiologic techniques have been useful for defining potential markets, determining how a drug is actually being used in the population, and determining unmet medical and public health needs. Further, the methods of epidemiology are useful for studying high-risk groups such as the elderly, the poor, and expectant mothers, thus providing important knowledge about the relative benefits and risks of therapy in populations rarely studied in clinical trials.

Estimating the incidence and prevalence of a disease is crucial for evaluating the current and projected future unmet medical need for drugs in development. These epidemiologic data provide critical information for decisions about which drug candidates to develop, since the potential market of a drug is an important consideration in drug and portfolio planning. This is especially relevant given that drug development takes on average eight and a half years and costs an average of \$850 million. Of 10,000 screened compounds, only 250 enter preclinical testing, 5 clinical testing, and one is approved by the FDA. Successful companies thus must carefully choose which early candidates in their pipeline to pursue. Information regarding the descriptive epidemiology of a condition may lead to decisions to usher a candidate drug on a “fast track” or to apply for approval under the US or EU orphan drug legislation.

Epidemiologic studies of prevalence, the natural history of a condition, or the frequency with which complications of a condition occur are particularly important for portfolio planning long-term. The rich data resources available for public use from the US National Center for Health Statistics, the Agency for Healthcare Policy and Research, the National Institutes of Health, and similar agencies outside of the US, such as the Office of National Statistics in the UK, can be used for these studies, or alternatively this information may be derived from population-based studies commissioned by industry using primary or secondary data sources, although the cost and time investment is considerably higher. Epidemiologic studies are also used to better understand regional prevalence and incidence of disease, especially in emerging markets where the burden of disease is often poorly characterized. In addition to prevalence, epidemiologic studies can estimate the burden (cost and disability) associated with specific conditions, providing data helpful for valuing a drug to patients and society.

Patterns of Medication Use and Beneficial Effects of Drugs

Once a drug is approved, epidemiologic methods can be used to monitor its use and the type of patients who receive

the drug. Observational studies may be informative as to the frequency of off-label use and use of multiple medications. Epidemiologic methods and databases such as the National Ambulatory Medical Care Survey may be used to monitor trends in the prescribing of certain pharmaceutical products, changes in the characteristics of users over time, or to study patterns of medication use among high-risk populations, such as the elderly with cognitive impairment. These methods are also used to quantify the beneficial effects of drugs (see Chapter 21). Study endpoints may vary from outcomes such as well-being or quality-of-life (see Chapter 23) to more quantitative variables such as blood pressure level, direct and/or indirect cost savings, and utilization of the health care system (see Chapter 22). Post-approval studies of benefit are particularly relevant when the clinical trials have focused on surrogate measures of efficacy, and there is a desire for further information regarding a medication’s impact on mortality or other long-term health outcomes, as discussed in the first section of this chapter.

ISSUES IN PHARMACOEPIDEMOLOGY

Resources for Pharmacoepidemiology

To respond rapidly and responsibly to safety issues, high-quality 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 Chapters 11 and 12). These databases offer several advantages over *ad hoc* 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 is advantageous as various data sources are necessary for research in pharmacoepidemiology. The limitations of many automated data sets are well established and need to be considered before conducting a study on a newly marketed medication (see Chapters 11, 12, and 14). 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. A common research limitation of automated data sources is that sufficient numbers of users may not yet be recorded, or the medication may not be marketed in the country where the database is located. Some data resources suffer from a considerable "lag-time" between data entry and availability for research purposes. Further, even though many health maintenance organizations 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 not contained in the HMO's research database. For some databases, medical record review may not be feasible due to concerns about patient confidentiality or anonymity, especially following recently enacted legislation on the privacy of health records (HIPPA; see Chapter 19). Continuing study of the research validity of these databases is crucial, and should be pursued when feasible. Finally, results from these sources are often limited in their generalizability.

Interpreting Pharmacoepidemiology Studies

Careful, scientifically sound research carried out using resources of high quality does not guarantee that a study's findings will be appropriately interpreted. Assuming that a safety issue has been suitably addressed using appropriate data resources, the study findings may still be improperly interpreted or misused. This has implications for the pharmaceutical industry and patients who lose access to beneficial and safe medications. Regulatory agencies are also affected by having to devote scarce resources to evaluating erroneous safety issues in order to make regulatory decisions, and by the impact these interpretations have on the public's confidence in the regulatory process. Ultimately, a disservice has been foisted on the public by generating unwarranted fears, removing safe and effective drugs, and charging higher prices for pharmaceuticals.

The media misinterpretation of epidemiologic results may have an impact on drugs in the market, resulting in useful drugs being precipitously withdrawn from the market. Misinterpretation of epidemiologic studies perpetuates the impression that the discipline is weak by generating controversy over study results, while promoting needless anxiety on the part of both patients and physicians. In such circumstances, the weaknesses of these studies are emphasized and the strength of the discipline overlooked. As a result, the information that epidemiology contributes may be considered

of questionable usefulness. Greater understanding of the strengths and limitations of epidemiology is needed by the public, the media, the government, and often by industry itself. These diverse groups have common interests and, through their joint efforts, the discipline of pharmacoepidemiology may be improved by focusing support, assessing study quality, and advancing a greater understanding of the field.

The relationship between science and industry also contributes to the misinterpretation of research results, and causes distrust of pharmaceutical companies, and increasingly the academic institutions with which they partner. Academic-industry partnerships have been in place since the early twentieth century, but their primary purpose initially was developing research capabilities within the emerging pharmaceutical industry. After the World War II, and due in large part to the government's funding of biomedical research through the National Institutes of Health (NIH), these relationships declined. The 1980s witnessed a resurgence of industry funding with the flattening of growth in the NIH budget and the passage of the Bayh-Dole Act in 1980. The academic-industry relationships that followed have clearly resulted in benefits to society, particularly more timely and effective technology transfer, but are plagued by concerns about researcher bias and the failure to communicate results adequately or, in some cases, at all.

Academic institutions, the NIH, and companies have responded to these concerns in multiple ways. However, the rules in place and processes used to manage potential conflicts of interest vary significantly. It is now generally recognized that there is a need for disclosure of financial interest, and often a limit on researchers having significant financial interests in a company that supports their research. In the future, conflicts of interest other than financial should also be examined, including peer or group recognition, career advancement, or political affiliations. For clinical research, it is also essential to ensure that appropriate processes are in place to safeguard research participants and their confidentiality. Detailed recommendations on these processes, such as when and how to convene Data Safety Monitoring Boards or guidelines for conducting research in pharmacoepidemiology, are currently available. Further guidance is needed to assist companies, universities, and regulatory agencies in defining types of conflict of interest, particularly non-financial ones, and in clarifying when these conflicts are significant and reportable.

CONCLUSIONS

In conclusion, epidemiology makes a significant contribution to the development and marketing of safe and effective

pharmaceutical products worldwide. It facilitates the regulatory process and provides a rational basis for drug safety evaluation, particularly in the post-approval phase. 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. Achieving this goal requires financial and intellectual support as well as a better understanding of the nature of the discipline and its uses. The need for scientists with training and research experience in pharmacoepidemiology has never been greater. Epidemiologists within industry have an opportunity to build on the successes of the last 20 years by advancing the methods of drug safety evaluation and risk management, and applying epidemiologic designs and methods to new areas within industry.

A VIEW FROM REGULATORY AGENCIES

KEY PRINCIPLES

The protection of public health is central to decision making by pharmaceutical regulators. Given this public health-oriented approach, several key concepts underscore the regulatory process.

Regulators have obligations to ensure that medicines on the market are of acceptable safety, quality, and efficacy. We approach this by taking evidence-based decisions and balancing risk and benefits from a population perspective at different stages of the product life cycle. Pharmacoepidemiology can make important contributions to these decisions, which impact a wide range of people, particularly the end-users of medicines. Regulators also need to respond when study results of potential public health importance are published, adopting critical evaluation and taking any regulatory action if necessary.

When making these decisions, a wide range of study designs, broadly divided into descriptive and analytic studies, are employed. The concept of an “evidence hierarchy” based on study designs is helpful. The hierarchy reflects the robustness of the data available. Descriptive studies, which include single case reports, case series, and uncontrolled cohorts or registries, limit the inferences that we make about causality. While largely hypothesis generating, these studies are most frequently the basis of post-licensing regulatory actions when time and resources do not permit more thorough analytic studies. Analytic studies include comparators and have the ability to test specific hypotheses. These include case-control studies, cohort studies, nested

case-control studies, and case-crossover analyses, as well as randomized clinical trials. It is accepted that randomized controlled trials offer greater control of bias in the study design and are at a higher level in the evidence hierarchy than observational analytical studies. Meta-analyses offer methods of amalgamating data scientifically from different studies.

When we consider the safety of medicines, we have a complex interaction of regulatory medicine, medicines’ safety, and pharmacoepidemiology. This interaction can be better understood by considering three axes (see Figure 6.1). While it is clear that safety concerns can arise at any time in the medicine’s life cycle (axis 1), data elements from different sources in the evidence hierarchy (axis 2) fulfill different roles in the evolution of safety concerns (axis 3), from generation of data on safety, to evaluation, to hypothesis testing.

In the context of a major safety concern, observational, descriptive, and analytic studies are only a subset of all the available data that should be considered for relevance. In addition to these data sources there is a wealth of other data that should be considered in an overall assessment of a medicine safety issue. These include data based on pharmacodynamic and pharmacokinetic studies (see Chapter 4) and nonclinical studies, including *in vitro* and animal studies.

From the public health perspective, there are several key concerns in relation to the nature and use of pharmacoepidemiologic data. These include the nature of the hypothesis, the aims and objectives posed, methodology details, ethical considerations, and data quality. The Excellence in Pharmacovigilance model of the UK Medicines and Healthcare Products Regulatory Agency (MHRA) outlines two main global aims in pharmacovigilance: the detection of harm and the demonstration of safety. The first is heavily dependent on spontaneous adverse drug reaction (ADR) reporting systems to flag previously unrecognized rare safety signals, including unusual patterns or excessive numbers of anticipated risks (see also Chapters 7 and 8). To demonstrate safety post-licensing, the evaluation of safety requires planned collection of outcome and exposure data on a sample of the newly exposed population until preset milestones of patient exposure are met.

Pharmacoepidemiologic data are used to ensure the maximum benefit at the minimum risk for the end-user of the medicine. The goal of analyzing spontaneous ADR reports is to maximize the detection of unrecognized safety issues and to minimize the chance of missing a safety signal. The patient and public health perspective is central to assessment of the impact of a given report of a suspected ADR on the medicine’s benefit/risk profile.

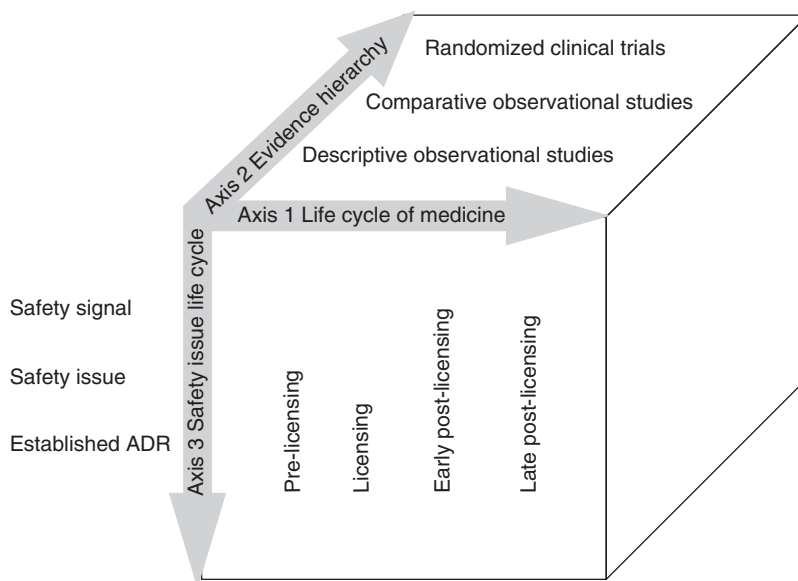


Figure 6.1. Jane's cube: The interaction of regulatory medicine, medicines' safety, and pharmacoepidemiology.

The highest professional standards must be applied in the design and conduct of post-licensing studies. A protocol with clear objectives, an independent advisory committee, and ethical review will help ensure that these studies generate useful safety data and dispel the concern of patients and health professionals that some post-licensing safety studies are principally for promotional purposes.

Regulation and Ethics in Research

In the US, the FDA has published guidelines for industry entitled "Good Clinical Practice: Consolidated Guidance." Good clinical practice (GCP) is an international ethical and scientific quality standard for designing, conducting, recording, and reporting trials that involve the participation of human subjects, applicable to both clinical trials and post-marketing studies. Compliance with this standard provides public assurance that the rights, safety, and well being of trial subjects are protected, consistent with the principles that have their origin in the Declaration of Helsinki, and that the clinical trial data are credible.

For prospectively conducted post-licensing studies, regulations require that the highest possible standards of professional conduct and confidentiality must always be maintained and any relevant national legislation on data protection should be followed (see also Chapter 19). The patient's right to confidentiality is paramount. The patient's

identity in the study documents should be codified, and only authorized persons should have access to identifiable personal details if data verification procedures demand inspection of such details. Identifiable personal details must always be kept in confidence. In the US, the Belmont Report is the basic foundation on which current standards for the protection of human subjects rest. Much of the biomedical research conducted in the US is governed either by the rule entitled "Federal Policy for the Protection of Human Subjects" (also known as the "Common Rule," which is codified for HHS at subpart A of Title 45 CFR Part 46), and/or the FDA Protection of Human Subjects Regulations at 21 CFR Parts 50 and 56. Research conducted on existing medical records must also consider data protection, anonymization, consent, and confidentiality. In the US, researchers conducting retrospective reviews of medical records must also take steps to ensure patient privacy and the protection of associated medical records.

DRUG LIFE CYCLE

Pre-licensing

Epidemiology Informs Key Milestones in Medicines Development

In drug regulation, pharmacoepidemiology has, to date, been most used as an aid to pharmacovigilance. However, there are also numerous applications of epidemiological

methodologies long before a medicine is licensed and used in the marketplace. When a pharmaceutical company is selecting potential disease targets to pursue and at key milestones during medicines development, epidemiologic techniques can be used to estimate and measure potential market size, the demographics of the diseased population, unmet medical needs, and how existing therapies are used in treatment. Such techniques can also be applied to further evaluate risk factors associated with adverse events observed during this period.

During medicines development, the traditional way to learn about the safety of a product is through the systematic collection of adverse event data during randomized, comparative clinical trials. However, epidemiologic techniques, including more descriptive methods, can supplement clinical trial data. For example, after the randomized period of a clinical trial, patients often continue on study therapy in an unblinded “extension phase.” Although the adverse event data are less robust than those from the randomized study, they provide useful additional information on the product’s safety, including valuable long-term exposure data. Similar descriptive safety analyses can be conducted when patients receive an investigational medicine on a “compassionate use” or “named patient” basis. Indeed, some regulatory authorities will only allow such use if protocols are put in place for the systematic collection of adverse event data. In the global medicines market, a medicine may be investigational in one country or region and already licensed and marketed in others. In this situation, safety data, such as spontaneously reported ADRs originating from the region where the medicine is marketed, can supplement the randomized data from the region where the medicine remains investigational. When adverse events are observed during clinical trials, epidemiologic techniques, such as nested case-control studies, can be used to understand better the risk factors associated with the adverse events. Such information can inform companies and regulators about populations at risk that can be used to more effectively manage risks post-licensing. Finally, while randomized studies are conducted, epidemiologic studies of the disease being treated and its existing therapy can be conducted.

Epidemiologic techniques can also be used to collect and analyze efficacy data (see Chapter 21). In some situations, for example when a disease is very rare or when conducting comparator trials may be judged unethical, the only possible way to collect efficacy data may be through epidemiologic techniques.

When considering the role of pharmacoepidemiology in the assessment of a medicine’s efficacy and safety, the

importance of confounding by the indication must always be borne in mind in the analysis and interpretation of such studies (see Chapters 16 and 21). When comparing treated patients with untreated patients, treated patients will have a higher rate of any disease that the medicine is intended to treat, although studies of the medicine’s effectiveness may be considered in some situations where effects are so dramatic that no comparator group is required. Randomization operates to control confounding in the study of intended effects.

Safety Assessment

To be licensed, the balance of benefits and risks of a medicine is judged acceptable for the indications granted. However, regulators are often questioned on how or why major medicine safety issues arise subsequently. To understand why our knowledge of safety at licensing is provisional, it is important to consider the extent and nature of the pre-licensing drug safety assessment, the limitations of clinical trials, and situations where a more extensive safety database may be needed.

Individual clinical trials are generally powered to answer specific efficacy questions with tight inclusion and exclusion criteria and they are usually of limited duration. Although the rate of common ADRs can also be estimated, such trials are unlikely to observe rare ADRs or reactions that only follow longer-term exposure.

Assessment of clinical trial safety data must be undertaken with the aim of minimizing the risk to future trial participants and patients. This assessment should take carefully considered case definitions and time dependency into account. A single case report of a suspected unexpected serious adverse drug reaction (SUSAR) from a clinical trial may prompt the use of analytic tools such as data mining and other signal detection strategies on these data (see Chapters 7 and 8).

Good clinical risk assessment depends on adequately designed and conducted preclinical studies, clinical pharmacology studies, and clinical trials programs to ensure that sufficient safety data are generated to allow for licensing of the product. The size of the pre-authorization human safety database needed depends on many factors, including the product, the population, the indication, the duration of use of the drug, and the results of the preclinical and clinical pharmacology programs.

Safety data, ideally, should be comparator-controlled safety data, including long-term safety data, to allow for comparisons of event rates and for accurate attribution of adverse events. Data should be available extending over a range of doses and in a diverse population. Risk assessment

should address potential interactions (drug–food and drug–drug interactions), demographic subpopulations, and effects of comorbid diseases.

International Conference on Harmonisation (ICH) guideline E1 outlines the size of the human database needed for licensing a medicine for non-life-threatening conditions. It recommends that data on at least 1500 patients be available when chronic/recurrent treatments for non-life-threatening diseases are considered, with 300–600 exposed for more than 6 months and 100 for more than 12 months.

The Role of Scientific (Regulatory) Advice During a Medicine's Development

In both the EU and the US, systems exist for pharmaceutical companies to obtain scientific and regulatory advice during their development of a medicine. In the US, the FDA encourages frequent interactions and scientific dialog with application sponsors throughout a product's life cycle. When planning pharmacoepidemiology studies, seeking the best advice from regulators and outside experts in epidemiology should improve the quality of the study and make clear the outcomes and expectations of such an endeavor.

The Relevance of Pharmacoepidemiology to Orphan Medicines

An orphan disease is a rare disease and an orphan medicine is therefore (logically) a medicine to diagnose, prevent, or treat a rare disease. The 1983 US Orphan Drug Act guarantees the developer of an orphan-designated product several incentives, including 7 years of market exclusivity following US market approval in the same indication.

Pharmacoepidemiology plays a central role in the consideration of orphan medicines, first in designating a medicine as an orphan product and second in supporting the collection of data demonstrating the safety and efficacy of the product needed to license it. In the US, a rare disease or condition is defined as any disease or condition that affects fewer than 200 000 people in the US. In applying for orphan designation, companies have to substantiate that the disease to be diagnosed, prevented, or treated has a prevalence below the legal threshold. Such substantiation usually requires the application of pharmacoepidemiologic techniques. For example, a company could use a longitudinal patient database in a locality where they are trying to establish prevalence and search the database for all cases of a particular condition. The number of cases in the entire country or region can then be calculated if one knows the total population covered by the database and the total population of the country or region.

The three pillars of drug regulation are quality, safety, and efficacy. Epidemiology may have a role to play in establishing the safety and efficacy of orphan medicines. The licensing criteria applied to orphan medicines are the same as those applied to any other medicine. However, the rarity of patients and their dispersal over a large area may make the conduct of a randomized study impractical, or even impossible. Randomized, parallel, double-blind controlled trials may be extremely difficult, especially if there is no current treatment available and the condition is life-threatening. If such a study design is not possible then alternative designs will have to be employed and epidemiology may have a key role to play.

Licensing a Medicine

Before a medicine can be marketed it is necessary to obtain a product license. The licensing of a medicine is a key step in a product's life and the licensing system is the main tool that regulators have to protect public health, ensuring that only medicines meeting strict criteria of quality, safety, and efficacy reach the market. To obtain a product license, pharmaceutical companies have to submit detailed documentation relating to the product and its development.

Regulatory authorities assess applications and make decisions on whether a medicine can be licensed. As well as making decisions on the overall balance of benefits and risks of the medicine (and therefore whether a license can be granted), the regulatory authorities must also make decisions on how the product should be used in the marketplace, including its indications and contraindications for use. The license includes regulated information about the product aimed at the medicine's users.

The ICH E2E guideline "Pharmacovigilance Planning" (2003) provides a structured method for summarizing the risks associated with a drug and for presenting a pharmacovigilance plan for when the product is marketed. The guideline is intended to aid industry and regulators in planning pharmacovigilance activities, especially in preparation for the early postmarketing period of a new medicine. The ICH guideline uses the term "safety specification," first for a document presenting the identified risks of a medicine, the potential for important unidentified risks, and the potentially at-risk populations and situations that have not been studied pre-licensing. The safety specification is intended to help identify the need for specific data collection in the post-licensing period and also to facilitate the construction of the pharmacovigilance plan. The pharmacovigilance plan is based on the safety specification. It sets out the proposed methods for monitoring the safety of the product, including

both “routine pharmacovigilance,” i.e., the methodologies such as spontaneous reporting and periodic safety update reports that are required of companies by law, and any specific studies planned as a result of risks or potential risks identified in the safety specification.

Whereas the ICH guideline “Pharmacovigilance Planning” provides a structured method for documenting the risk profile of the product and planned safety monitoring, it does not deal with how to minimize risks to patients (other than through effective safety monitoring). Regulatory authorities are encouraging and in some instances requiring that measures to minimize the risks to patients from the product are documented, assessed, and agreed on during the licensing process. Here again, the epidemiologist can play a central role.

Post-licensing

At the time of licensing, due to the limitations of clinical trials in simulating the complexities of “real-world” use, we generally have incomplete knowledge about the safety of a new medicine. For most medicines, following launch onto the market, the exposure to a medicine increases from a few hundred or thousands of patients exposed during the development program, to tens or hundreds of thousands or even millions of patients. With the increasing globalization of the pharmaceutical industry, this mass exposure can occur within months of a product launch. Furthermore, the controlled way the medicine was used during development switches to the relative anarchy of everyday prescribing, dispensing, and usage of medicines. With the general availability of the product, we learn about the effects of a medicine in everyday practice, including rare ADRs and ADRs that only occur after prolonged use, as well as ADRs associated with co-prescribing with other medications and those unique to or enhanced by comorbidities in the treated population. The additional knowledge of the safety profile in normal clinical use must be systematically managed and evaluated for the protection of patients.

The epidemiologist’s role is much better established in the post-licensing phase of a medicine’s life: the epidemiologist plays a central role in pharmacovigilance, but may also be involved in the variation, renewal, and reclassification of medicines. In addition, governments are increasingly requiring data on cost-effectiveness and relative effectiveness prior to including a new medicine in formularies for use or prior to agreeing to reimburse patients for the cost of the medicine (see Chapter 22). Here again the epidemiologist may play a role in the collection, analysis, or presentation of data.

Pharmacovigilance

The monitoring of the safety of marketed medicines is known as pharmacovigilance, defined by the WHO as “the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug related problems.” The subsequent sections describe the practice of pharmacovigilance.

Data Collection and Management

Data on drug safety from all available sources need to be collected and managed systematically to be able to detect possible drug safety hazards as effectively as possible.

Pharmaceutical companies have obligations to collect all data relevant to the safety of their products and submit such data to regulators in line with guidance and legislation. Regulators monitor these data for signals but also collect and screen safety data on medicinal products for signal detection independently of pharmaceutical companies. Safety data collection is carried out throughout the post-licensing lifetime of the product, until the product is discontinued or withdrawn, as new safety issues can emerge and have emerged at any time, even with well-established products.

The collection and management of data has to be systematic, incorporating quality assurance and control measures, utilizing necessary resources, skills, and equipment to ensure timely access to the data for signal detection. Widely agreed upon definitions, standards, contents, and conditions for case reporting, including for electronic transmission for individual case reports, have been established. There are also agreed standards, content, and formats for periodic safety update reports (PSURs) for companies which are submitted to regulators at fixed time points from licensing. Other safety data and potential safety signals will come to the attention of regulators through processes involving applications to change product licenses (variations), post-licensing commitments and follow-up measures (agreed at the time of licensing), regular screening of the published literature, communication among regulators, and patient and health professional enquiries.

Signal Detection

A signal is defined by the WHO as “reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously. Usually more than a single report is required to generate a signal, depending upon the seriousness of the event and the quality of the information.” Historically, most medicine safety signals have come from spontaneously reported suspected ADRs. However, major

safety issues may be detected from any of the data sources relevant to a medicine's safety.

The signal detection methodologies in common use by regulatory agencies based on spontaneous reports of suspected ADRs must be considered in the context of the strengths and weaknesses of such a monitoring system. Used to generate information about rare and previously unknown ADRs, spontaneous reports are collected largely through passive surveillance systems where reporting of suspected ADRs to regulatory authorities is voluntary for health professionals (in most countries) but statutory for license holders. In some countries, notably the US, suspected ADR reports are also accepted directly from patients.

Spontaneous ADR reports are most useful where the reaction is unusual and unexpected in the indication being treated and where the ADRs occur in a close temporal relationship with the start of treatment or following a dose increment. From the regulatory viewpoint, these are reports of *suspected* ADRs and the unique feature of spontaneous reporting systems is that the suspicion of the reporter has been captured. An assessment of individual ADR reports may indicate whether there could be alternative explanations for the observed reaction other than the medicine. Poor quality and/or incomplete information in the case report often makes interpretation of the causal relationship between the product and the observed reaction, as well as wider generalization, difficult. The number of ADR cases reported may not be a good indicator of a signal as channeling of high-risk patients to newer therapies also leads to increased reporting with a newer agent. ADRs are less likely to be suspected and reported spontaneously where the reaction has an insidious onset, the reactions occur only following long-term treatment, or where the disease being treated has a high incidence of similar outcomes. In addition, those ADRs that are caused by a lack of efficacy may not be considered as ADRs and therefore not reported. Underreporting is a feature of all ADR reporting systems. The frequency of reporting for a given medicine varies over time, with time from first marketing, and with periods of media activity surrounding the product. Given the variability in reporting and the numerous factors that affect reporting, it is well accepted that reporting rates cannot be used to reliably estimate incidence rates and that comparison of reporting rates between medicines or countries may not be reliable.

Generally, spontaneous ADR reports are examined by systematic manual review of every report received. As an aid to signal detection, screening algorithms based on automated signal detection systems have been explored. Such methods have been referred to as data mining. The aim of these statistical aids is to provide the means of comparing

the frequency of a medicine–event combination with all other such combinations in the database under consideration, with the potential for early detection of signals of potential medicine–event associations (see Chapter 8). Any such signals must be confirmed by detailed evaluation by skilled clinicians and epidemiologists of the case reports that generated the signal.

With data mining, signals are generated without external exposure data, adverse event background information, or medical information on ADRs. Further detailed evaluation of relevant data is needed. The systems cannot distinguish between already known associations and new associations, so the reviewers must filter these known reactions.

Prioritization: Impact Analysis Concepts

Detailed signal evaluation using all the relevant data is complex and resource intensive. Regulators therefore need to prioritize signals. The potential impact of a safety issue on public health is the foundation for regulators' prioritization but, to date, the judgment of impact has been based on qualitative and subjective criteria. Such criteria include "SNIP": regulators would prioritize relatively Strong signals that are judged to be New, clinically Important, and have the potential for Prevention.

Signal Evaluation and Risk Quantification

The initial steps of evaluation of a potential medicine safety issue will focus on causality assessment, identification of any other possible causes of the adverse events being reported, and assessing the risk to both individuals and the public, in terms of both frequency and seriousness of the reactions.

When a signal of a suspected ADR arises from spontaneous reports, any other similar cases reported previously, forming a case series, should also be evaluated. Reporting rates—the number of ADR reports received divided by the estimated exposure to the product—can be useful for hypothesis generation. However, they are subject to many limitations. Numerators are subject to known variability and under-ascertainment. The choice of denominator should be dictated by the medicine safety question: for example, all use of a particular route of administration or all use in children.

Ideally, systematic studies should be reviewed in order to estimate the incidence of an ADR and the confidence interval around the estimate. The calculation of frequency estimates using patient-time as the denominator assumes that the rate of a hazard is constant. However, three models of hazard function (the instantaneous incidence rate) can be summarized. In the peak-shaped hazard model, the hazard

increases rapidly over an initial period and then drops to baseline level (e.g., clozapine-induced agranulocytosis). In the constant hazard model the rate reaches a plateau shortly after the beginning of treatment (e.g., upper gastrointestinal bleeding associated with NSAIDs). Lastly, in the increasing hazard rate model the hazard continues to increase over time (e.g., hormone replacement therapy and breast cancer). Additionally, using the upper one-sided 95% confidence interval reflects the uncertainty in the data and gives a worst-case scenario.

Evaluation of spontaneous reports should consider demographic factors such as age, gender, race, or other subgroups, the effects of exposure dose, duration, the effects of time, the effects of other drugs, comorbid conditions, and/or the target population. Owing to the nature of the data, spontaneous ADR reports generally do not permit a direct conclusion on the association between a particular ADR and the medicine in the population. However, some factors in case reports that strengthen the association include the presence of a positive re-challenge, positive de-challenge, and a clear absence of alternative causes. Classifications have been derived for assessment of the likelihood of causality in individual spontaneous reports (see Chapter 17). If a safety issue warrants detailed assessment, the evaluation must be widened to consider all available data, including preclinical, clinical pharmacology, clinical trials, pharmacoepidemiology studies, and class effects.

The Assessment of Non-spontaneous Data

Data employed in a regulatory risk assessment are critically reviewed, bearing in mind limitations of data derived from sources at different levels of the evidence hierarchy. The pharmacoepidemiologic data sources (other than spontaneous reports), include: active surveillance, registries, comparative observational studies, clinical trials, large simple safety trials and meta-analyses.

Benefit/Risk Assessment

Assessment of the balance of benefit and risk is conducted throughout the life of a medicine, through medicine development, at the time of license application, and then continuously in the post-licensing phase. The principles underlying benefit/risk assessment are the same at all stages. However, the data that may be available will differ substantially. Every year products are withdrawn from various markets around the world for safety reasons. However, regulators and companies have often used different methods of benefit/risk assessment and reached very different conclusions. The lack of standardization of benefit/risk assessment led the Council for International Organizations of Medical

Sciences (CIOMS) to set up a working group to produce guidelines on standardized benefit/risk assessment.

In the post-licensing phase, once a safety issue has been identified, and evaluation has resulted in a judgment that the medicine may be a significant threat to public health, it is important to conduct a thorough benefit/risk assessment. Benefit/risk assessment can be made robust by following a systematic plan and ensuring that all relevant data are considered.

The *disease being treated* by the medicine under investigation will have a major impact on the balance of benefits and risks. For example, if the disease is self-limited, serious ADRs would have a considerable negative impact on the balance of benefits and risks. In contrast, with a disease with a high mortality, serious ADRs may still be outweighed by the benefits afforded by the drug. In addition to describing the natural history of the disease, it is also important to describe the demography of the disease, including its incidence and prevalence. This allows all the patient groups likely to be exposed to be considered, as well as the public health impact of benefits and risks to be judged.

The *population being treated* also needs to be considered. When used in normal clinical practice, a medicine may not be used within the confines of the licensed indication. But differences may occur in how medicines are handled by different populations.

Just as the nature of the disease being treated is important, the *purpose of the intervention* should also be described. Medicines may be used for prevention, treatment, as part of a procedure, or for diagnosis. These factors will impact the balance of benefits and risks. For example, a medicine used to prevent disease in an otherwise healthy individual must have a very well-established safety profile, particularly if the disease being prevented is rare or non-serious.

The *therapeutic alternatives* to the medicine being evaluated should be identified. For the majority of diseases there will be alternative medicines available. Once the main alternatives to the treatments under evaluation have been identified, their benefit and risk profiles should be considered.

For most medicines, an *evaluation of benefits* in normal clinical use has not been conducted and therefore clinical trial efficacy has to be used as a “surrogate” of benefit. The robustness of premarketing efficacy data is often far superior to that available for risk, as clinical trials are designed first and foremost to demonstrate the efficacy of the product. Another important consideration when assessing benefit is whether efficacy has been demonstrated in terms of clinical outcomes. For example, when many of the antiretroviral agents were licensed to treat HIV infection and

AIDS, only surrogate markers of clinical endpoints were available, such as increases in CD4 lymphocyte count and reduction of HIV RNA viral load. Only subsequently have morbidity and mortality data confirmed the major benefit of these medicines in the treatment of HIV-infected individuals. When a major safety issue occurs for a medicine where robust data on clinical benefit are available, the judgment on the balance of benefits and risks might be different from that for a medicine where only data using surrogates of clinical endpoints exist.

The longevity of the effect of a medicine should also be taken into account. For chronic disease, the initial efficacy may be marked, but if this is not sustained the overall benefit of the medicine may be very limited. This may be due to tachyphylaxis or, for infectious diseases, the development of resistance.

Another factor to consider is whether the medicine being evaluated is to be used as *first or second line therapy*. For example, a medicine that is used as a first line chemotherapeutic agent where alternatives exist might be judged to have a different balance of benefits and risks from a chemotherapeutic agent which is only indicated in patients who had failed all other treatments. This concept of first and second line therapy is often used when restricting the use of medicines with a major safety issue.

The *degree of efficacy* needs to be documented for each indication (whether licensed or not) and each population treated. For example, ACE inhibitors are indicated for post-myocardial infarction prophylaxis, cardiac failure, hypertension, and to prevent renal damage in patients with diabetes mellitus. Convincing mortality data exist showing a benefit of ACE inhibitors in some but not all of these indications. Therefore if a major safety issue arose, the balance of benefits and risks might be different for the different indications.

Once causality has been established between the adverse event and the medicinal product, assessment of its seriousness and severity will help in the judgment of the *impact of the risk* on the individual and population being considered. In order to identify ways to reduce risk, it is essential to try to investigate whether the risk is associated with a particular patient group, particular dosage, results from an interaction, or whether there is an early warning sign.

The *frequency of the adverse reaction* must be assessed. Frequency will impact on the benefit/risk balance for an individual, as well as the impact of the medicine's toxicity on populations. It is also important to consider the overall *adverse reaction profile* of the medicine.

As well as documenting what the alternative therapies are, it is helpful when assessing risks to select one comparator medicine, possibly of the same class and indication,

for a direct head-to-head comparison. If a comparator can be selected then both the benefits and risks should be compared between the two medicines in all indications and populations.

Having described the target disease and population being treated, the purpose of the intervention and alternative therapies, and evaluated the benefits and risks, it is then necessary to try to judge the *overall balance of risks and benefits* of the medicine. Benefits and risks for all indications and populations need to be taken into consideration. The overall balance may be very difficult to judge: the type of evidence available for benefit may be very different from that available for risk. The concepts of *number needed to treat* for benefit and *number needed to harm* for risk can help to quantify and may therefore aid in the comparison. Whenever possible, some estimate of the number of serious ADRs that would occur for a given positive outcome should be attempted.

Peer review of any assessment is an important step in ensuring its quality, and involving a range of additional experts in the judgment should result in a more balanced decision.

Case Example 6.3 provides an example of a recent benefit/risk assessment.

CASE EXAMPLE 6.3: SAFETY OF SELECTIVE SEROTONIN REUPTAKE INHIBITOR (SSRI) ANTIDEPRESSANTS

Background

An Expert Working Group of the UK Committee on Safety of Medicines (CSM) was convened to investigate ongoing public concern about the safety of SSRIs.

Issue

The Group considered:

- All available evidence in relation to any association of SSRIs with suicidal behavior, suicide attempts and behavioral disorders, withdrawal reactions or dependence, and any implications for the risk/benefit of these products against the background of the burden of depressive illness and epidemiology of suicide and non-fatal self-harm.
- The adequacy of the product information for SSRIs.
- Recommendations where necessary for changes in the product information and wider communications.

(Continued)

Approach

- Key data sources assessed in the review included published and unpublished clinical trials, spontaneous reports from patients and health professionals, evidence from experts, epidemiological studies on a primary care database, and reviews of published literature.
- Relevant companies were asked to supply all data from all randomized controlled clinical trials undertaken for their products.
- Each company was asked to analyze the results from all relevant trials to a pre-specified protocol and to supply case narratives for all reports of suicidal behavior. These data were checked by the UK Medicines and Healthcare Products Regulatory Agency (MHRA) for completeness and consistency.

Results

- Based on the work of the group, the CSM issued advice on the use of SSRIs in children and adolescents. The balance of risks and benefits for the treatment of depressive illness in under-18-year-olds was judged to be unfavorable for paroxetine, venlafaxine, sertraline, citalopram, and mirtazapine. It was not possible to assess the balance of risks and benefits for fluvoxamine due to the absence of data. Fluoxetine was considered to have an overall favorable balance of risks and benefits in treating depressive illness in children and adolescents.
- There was no evidence of an increased risk of self-harm or suicidal thoughts in young adults of 18 years and over in association with SSRIs. As a precautionary measure, however, it was recommended that young adults treated with SSRIs should be closely monitored.
- With adults, based on clinical trial data, a modest increase in risk of suicidal thoughts and self-harm in SSRI-treated compared with placebo-treated groups could not be ruled out. There were insufficient data from clinical trials to conclude that there was a marked difference among SSRI class members or between SSRIs and tricyclic antidepressants (TCAs). Epidemiology studies based on primary care databases indicated that there was no difference in risk of suicidal behavior between SSRIs and TCAs.
- The MHRA was aware of three recent studies of the association between antidepressants and suicidal behavior using the General Practice Research

Database (GPRD). These studies contain overlapping sets of patients but used different inclusion and exclusion criteria. The findings of the three studies were broadly consistent with no evidence of an increased risk of suicidal behavior in adults exposed to SSRIs compared to a range of other antidepressants. However, there was evidence that children and adolescents exposed to SSRIs were at an increased risk of suicidal behavior compared to those exposed to other antidepressants. The questions of confounding by indication, whereby patients thought to be at greater risk of suicidal behavior were preferentially prescribed SSRIs owing to SSRIs' relative lack of toxicity in overdose, was an important consideration in the interpretation of the results.

Strengths

- The burden of depressive illness and epidemiology of suicide and non-fatal self-harm provided essential background information, including definitions, risk factor group, and incidence and prevalence data.
- Advice to health professionals and patients was evidence-based and comprehensively covered all available data sources.
- Patients' perspectives were a key theme in considerations.
- Epidemiology studies based on computerized and anonymized clinical records from primary care are useful when the medicine is regularly used in this care setting.

Limitations

- The review was based on data available to date and represented a snap-shot in time.
- Epidemiology studies based on computerized and anonymized clinical records from primary care are not without limitations such as completeness of recording and use of prescribing as a proxy for exposure status when such medicines may not be taken.
- The need for further research on the use of antidepressants was identified, such as the effectiveness of SSRIs in mild depression, factors governing general practitioners' reasons for prescribing antidepressants, and the epidemiology of depressive illness in children.
- Further standardization and refinement of clinical trial methods for assessment of antidepressants were identified.

Summary Points

- Any drug taken to treat a medical condition has potential risks as well as benefits.
- Randomized controlled trials may detect common ADRs but they generally do not include sufficient people to detect rare adverse effects such as suicide.
- Post-authorization pharmacovigilance is essential, and each data source has its own strengths and limitations.
- A pro-active approach with risk management planning is the gold standard and this should include epidemiology of the target conditions and possible key risks.
- Epidemiology studies in medical record databases can provide a vital contribution to the evidence base of a risk/benefit review.
- Important new safety issues can arise even with well-established medicines.
- Pharmacovigilance involves: collecting information, managing that information, detecting safety signals, assessing those signals, benefit/risk assessment, taking action to reduce risk, communicating with stakeholders, and audit of the actions taken.
- Benefit/risk assessment should involve evaluation of the benefits and risks in all indications and populations, whether or not a particular use is licensed.
- Assessment of safety issues involves multiple players and the result will be more robust if the perspective of stakeholders is taken into consideration.

Action to Reduce Risk or Increase Benefit and Communication

Following the benefit/risk assessment it will usually be necessary to take action either to increase the benefit of the medicine by improving its rational use or to reduce the risk, which usually involves educating practitioners and consumers, but on occasion requires restricting its use.

The term “regulatory action” typically refers to action taken in relation to licenses. However, here we use a broader definition covering all measures that could be taken by regulators or companies to improve the benefit/risk balance. Regulatory action may be voluntary, where for example a company voluntarily submits a variation or cancels a license. In contrast, the regulatory authority may take compulsory action. Whenever possible the desired regulatory outcome should be achieved through voluntary action. Compulsory action is more likely to lead to litigation and should be reserved for major public health threats where agreement for voluntary action with the company cannot be reached.

When taking decisions on the appropriate regulatory action, certain guiding principles should be remembered:

- Objectivity: assessment and decision making should be evidence based and free from conflicts of interest.
- Equity: there should be equity of regulatory action on products if the risk assessment and particularly the benefit/risk assessment is the same for different products.
- Accountability: decision makers are accountable for decisions and regulatory action taken.
- Transparency: within the confines of regional laws on commercial confidentiality and data protection, decision making and action taken should be transparent to stakeholders.

If the overall balance of benefits and risks is judged to be negative, then the product will be withdrawn unless risk minimization strategies can be identified that would swing the balance away from risk or towards benefit (see Chapter 27).

Effective communication about safety issues is essential if ADRs are to be prevented. Communications about a safety issue need to be planned and a communication team will normally need to be established. The messages to be conveyed should be targeted, understandable, open, informative, and balanced. To ensure that messages are clear, concise, and understandable, it is wise to involve communications specialists in writing documents and, if time permits, to user-test messages before distribution. Different but compatible messages may be required for health care professionals, patients, and the media. Timing communications, particularly for urgent safety issues such as product withdrawals, is essential. If the media carry a major medicine safety issue that leads to patients consulting their health care professionals and those professionals have not been briefed in advance, then the regulators and pharmaceutical companies will be judged to have failed both professionals and patients. This can be a major challenge but must be the focus when planning safety communications. It is very easy for non-specialists to misquote and misrepresent data (including epidemiologic data) in communication documents, and the epidemiologist has a role in checking such documents for accuracy. In addition, by advising on data collection, including survey methods, the epidemiologist may have a role in verifying that information has been received, understood, and followed.

Audit (Measurement of the Outcome of Interventions)
Evaluation (audit) of the success or impact of regulatory action taken in response to a specific safety issue is an essential duty of companies and regulators to the public. The main

objective of most actions is to inform and change behavior, be it prescribing or dispensing behavior or the habits of the public. These objectives can be very difficult to achieve and this is why we cannot assume that our actions have been effective. The epidemiologist has a central role in measuring and judging the impact of regulatory interventions.

Other Regulatory Activities

In addition to pharmacovigilance, there are other roles for the epidemiologist in the post-licensing phase of a medicine's life. For many medicines, the initial licensed therapeutic use is just the first of an expanding range of uses. Further uses may be in different populations with the same disease, in populations with the same disease but of a different severity or at a different stage in the disease's evolution, in the same disease but as part of a different treatment regimen, or even to treat a completely different disease.

The epidemiologist may have a role in identifying potential new uses for the medicine. By studying the disease, the demographics of the populations affected, the current treatment, and natural history of the disease, additional potential uses may be identified. In addition to seeking expert opinion and understanding the pharmacology of the medicine, use of longitudinal patient databases can be very informative in identifying potential new uses. Just as with initial licensing, to obtain a licensed new indication, a company will need to obtain data supporting the safe and effective use of the medicine in the new disease area or population.

An additional regulatory measure is the control of the distribution of the medicine, including whether it can be obtained only with a prescription from a registered doctor, dispensed by a pharmacist without a prescription, or bought without the intervention of any health care professional. Companies need to apply to the regulatory authorities to change the prescription status; in some countries this is referred to as reclassification. Companies may wish

to have their medicine used without prescription, as this may increase sales, but only certain types of conditions, symptoms, or diseases are appropriate for non-prescription treatment. Disease factors include whether it is easily diagnosable without the intervention of a doctor and whether misdiagnosis could have serious consequences. The main factors related to the medicine itself are whether it is safe (and the safety profile is well established) and simple to take. It should also be remembered that the profile of the population receiving the medicine may change dramatically when the intervention of a doctor is excluded. The epidemiologist may play a role in establishing that a medicine can be used safely and effectively without the intervention of a doctor. If a medicine is reclassified to non-prescription use, new safety issues may emerge and renewed vigilance in safety monitoring will usually be required. This may be particularly challenging, as, in many countries, doctors remain the main reporters of suspected ADRs.

CONCLUSIONS

Pharmacoepidemiology has a central and increasingly recognized role in the regulation of medicines, its use underpinning more and more regulatory decisions. A major challenge ahead is improving the robustness and richness of the pharmacoepidemiologic data upon which decisions are based. The technical, scientific, and legal issues are challenging, including the need for rapid data access and analysis (for urgent safety issues), statistical power, dealing with bias and confounding, obtaining data from sectors of the health market where currently they are lacking, and consent and confidentiality (protecting the individual but not at the expense of harming the public). Despite these challenges, we believe the use of pharmacoepidemiology is making an important contribution to better regulation and better protection of public health.

Key Points

View from Academia

- Pharmacoepidemiology will play an increasingly prominent role in providing therapeutics knowledge as health care transactions are increasingly captured in computers, and these data can be manipulated by increasingly usable and powerful statistical packages.
- Beyond understanding the benefits and risks of therapeutics, better knowledge is needed about how to effectively apply technologies, drugs, biologics, diagnostic devices, and therapeutic devices, such that patients are likely to experience benefits greater than risks.
- At the local level, there is a pressing need for institutions and health systems to develop data repositories and local expertise in appropriate analysis related to quality. Academic Medical Centers (AMCs) have the expertise to set the example by organizing data repositories and providing access to data for the purposes of improving quality and developing a generalized understanding of diagnostic and therapeutic strategies.
- AMCs should seek to leverage their unique position of educator, health care provider, and researcher to move national practice towards higher quality health care.

- AMCs need to focus attention on translating medical research findings into statements that can be acted upon effectively by policy makers, practitioners, and the public, directly and through the media.
- A vital role of AMCs is to instill in practitioners a fundamental understanding of the principles of therapeutics and the measurement of quality in health care delivery; another is creative integration with health care providers, government agencies, and the broader medical industry.

View from Industry

- The safety profile of any drug reflects an evolving body of knowledge extending from preclinical investigations to the first use of the agent in humans and through the post-approval life cycle of the product.
- Results from clinical trials, spontaneous reports, epidemiologic studies, and, where relevant, preclinical data sets, should be evaluated for their potential to address safety questions, with close consideration given to the unique strengths and limitations of the study designs and data collection methods used.
- Epidemiology plays a central role in drug safety assessment and risk management activities within the pharmaceutical industry, whether through studies of the natural history of disease, disease progression/treatment pathways, and mortality and morbidity patterns, or in the design and implementation of post-approval safety studies or risk minimization programs.

View from Regulatory Agencies

- All medicines have potential risks, as well as benefits. Important new safety issues can arise even with well-established medicines.
- Post-authorization pharmacovigilance is essential, based on all available data sources, each data source having its own strengths and limitations. A proactive approach with risk management planning is the gold-standard.
- Pharmacovigilance involves: collecting information, managing that information, detecting safety signals, assessing those signals, benefit/risk assessment, taking action to reduce risk, communicating with stakeholders, and audit of the actions taken.
- Benefit/risk assessment should involve evaluation of the benefits and risks in all indications and populations, whether or not a particular use is licensed.

SUGGESTED FURTHER READINGS

ACADEMIA

- Blumenthal D. Academic-industrial relationships in the life sciences. *N Engl J Med* 2003; **349**: 2452-9.
- Butler J, Arbogast PG, Daugherty J, Jain MK, Ray WA, Griffin MR. Outpatient utilization of angiotensin-converting enzyme inhibitors among heart failure patients after hospital discharge. *J Am Coll Cardiol* 2004; **43**: 2036-43.
- Califf RM. The Centers for Education and Research on Therapeutics. The need for a national infrastructure to improve the rational use of therapeutics. *Pharmacoepidemiol Drug Saf* 2002; **11**: 319-27.
- Califf RM, DeMets DL. Principles from clinical trials relevant to clinical practice: Part I. *Circulation* 2002; **106**: 1015-21.
- Finkelstein JA, Davis RL, Dowell SF, Metlay JP, Soumerai SB, Rifas-Shiman SL *et al*. Reducing antibiotic use in children: a randomized trial in 12 practices. *Pediatrics* 2001; **108**: 1-7.
- Finkelstein JA, Stille C, Nordin J, Davis R, Raebel MA, Roblin D *et al*. Reduction in antibiotic use among US children, 1996-2000. *Pediatrics* 2003; **112**: 620-7.
- Food and Drug Administration. Innovation or stagnation? Challenge and opportunity on the critical path to new medi-

cal products. Available from: <http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html/>. Accessed: August 2, 2004.

Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press, 2001.

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**: 211-18.

Mebane F. The importance of news media in pharmaceutical risk communication: proceedings of a workshop. *Pharmacoepidemiol Drug Saf* 2004; **14**(5): 297-306.

Metlay JP, Shea JA, Asch DA. Antibiotic prescribing decisions of generalists and infectious disease specialists: thresholds for adopting new drug therapies. *Med Decis Making* 2002; **22**: 498-505.

Moynihan R. Drug company sponsorship of education could be replaced at a fraction of its cost. *BMJ* 2003; **326**: 1163.

Mudano A, Allison J, Hill J, Rothermel T, Saag K. Variations in glucocorticoid induced osteoporosis prevention in a managed care cohort. *J Rheumatol* 2001; **28**: 1298-305.

Mudano AS, Casebeer L, Patino F, Allison JJ, Weissman NW, Kiefe CI *et al*. Racial disparities in osteoporosis prevention in a managed care population. *South Med J* 2003; **96**: 445-51.

- Patino FG, Allison J, Olivieri J, Mudano A, Juarez L, Person S *et al.* The effects of physician specialty and patient comorbidities on the use and discontinuation of coxibs. *Arthritis Rheum* 2003; **49**: 293–9.
- Pearson SA, Ross-Degnan D, Payson A, Soumerai SB. Changing medication use in managed care: a critical review of the available evidence. *Am J Managed Care* 2003; **9**: 715–31.
- 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.
- Relman AS. Defending professional independence: ACCME's proposed new guidelines for commercial support of CME. *JAMA* 2003; **289**: 2418–20.
- Sung NS, Crowley WF Jr, Genel M, Salber P, Sandy L, Sherwood LM *et al.* Central challenges facing the national clinical research enterprise. *JAMA* 2003; **289**: 1278–87.
- Woolley RL. Centers for Education and Research in Therapeutics. *Clin Pharmacol Ther* 1994; **55**: 249–55.

INDUSTRY

- Blumenthal D. Academic–industrial relationships in the life sciences. *NEJM* 2003; **349**: 2452–9.
- 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.
- Food and Drug Administration. *Managing the Risk from Medical Product Use: Creating a Risk Management Framework*. Washington, DC: US FDA, 1999.
- Food and Drug Administration. *Guidance for Industry: Establishing Pregnancy Exposure Registries*. Rockville: FDA, 2002.
- Goldman SA. Limitations and strengths of spontaneous reports data. *Clin Ther* 1998; **20**: C40–4.
- Hauben M. A brief primer on automated signal detection. *Ann Pharmacother* 2003; **37**.
- Holmes LB. Teratogen update: *Bendectin*. *Teratology* 1983; **27**: 277–81.
- International Society for Pharmacoepidemiology (ISPE). 1996. *Guidelines for Good Epidemiology Practices for Drug, Device and Vaccine Research in the United States*. (<http://www.pharmacoepi.org/resources/goodprac.cfm>, Accessed August 5, 2004).
- Kaitin KI, Cairns C. The new drug approvals of 1999, 2000, and 2001: Drug development trends a decade after passage of the Prescription Drug User Fee Act of 1992. *Drug Inf J* 2003; **37**: 357–71.
- Lasagna L. The Halcion story: trial by media. *Lancet* 1980; **1**: 815–16.
- Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**: 915–20.
- Reiff-Eldridge R, Heffner CR, Ephross SA, Tennis PS, White AD, Andrews EB. Monitoring pregnancy outcomes after prenatal drug exposure through prospective pregnancy registries: a pharmaceutical company commitment. *Am J Obstet Gynecol* 2000; **182**: 159–63.
- Smalley W, Shatin S, Wysowski DK. Contraindicated use of cisapride: impact of Food and Drug Administration regulatory action. *JAMA* 2000; **284**: 3036–9.
- Weatherby LB, Walker AM, Fife D, Vervae P, Klausner MA. Contraindicated medications dispensed with cisapride: temporal trends in relation to the sending of “Dear Doctor” letters. *Pharmacoepidemiol Drug Saf* 2001; **10**: 211–18.

REGULATORY AGENCIES

- Current Challenges in Pharmacovigilance: Pragmatic Approaches*. Report of CIOMS Working Group V. Geneva: 2001.
- FDA Center for Drug Evaluation and Research Guidance for Industry: *Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment*, March 2005. <http://www.fda.gov/cder/guidance/6359OCC.pdf>.
- FDA Center for Drug Evaluation and Research Guidance for Industry: *Guidance for Industry Development and Use of Risk Minimization Action Plans*, March 2005. <http://www.fda.gov/cder/guidance/6358fnl.pdf>.
- Garcia Rodriguez LA, Perez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol* 1998; **45**: 419–25.
- Gargiullo PM, Kramarz P, DeStefano F, Chen RT. Principles of epidemiological research on drug effects. *Lancet* 1999; **353**: 501.
- Haffner ME. The current environment in orphan drug development. *Drug Inf J* 2003; **37**: 373–9.
- Henderson L, Yue QY, Bergquist C, Gerden B, Arlett P. St John's wort (*Hypericum perforatum*): drug interactions and clinical outcomes. *Br J Clin Pharmacol* 2002; **54**: 349–56.
- International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). E2E: Pharmacovigilance Planning. ICH, 2003. Available at: <http://www.ich.org>. Accessed March 2004.
- Jefferys DB, Leakey D, Lewis JA, Payne S, Rawlins MD. New active substances authorised in the United Kingdom between 1972 and 1994. *Br J Clin Pharmacol* 1998; **45**: 151–6.
- Jick H, Garcia Rodriguez LA, Perez-Gutthann S. Principles of epidemiological research on adverse and beneficial drug effects. *Lancet* 1998; **352**: 1767–70.
- McMahon AD, MacDonald TM. Design issues for drug epidemiology. *Br J Clin Pharmacol* 2000; **50**: 419–25.
- Rawlins MD. Spontaneous reporting of adverse drug reactions. II: Uses. *Br J Clin Pharmacol* 1988; **26**: 7–11.
- Rawlins MD, Jefferys DB. Study of United Kingdom product licence applications containing new active substances, 1987–9. *BMJ* 1991; **302**: 223–5.
- The Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, <http://www.who-umc.org>.
- Waller PC, Arlett P. Responding to signals. In: Mann R, Andrews E, eds, *Pharmacovigilance*. Chichester: John Wiley & Sons, 2002; pp. 105–28.
- Waller PC, Evans SJW. A model for the future conduct of pharmacovigilance. *Pharmacoepidemiol Drug Saf* 2003; **12**: 17–29.

SECTION II

SOURCES OF PHARMACOEPIDEMIOLOGY DATA

Spontaneous Reporting in the United States

The following individuals contributed to editing sections of this chapter:

SYED RIZWANUDDIN AHMAD,¹ NORMAN S. MARKS,² and ROGER A. GOETSCH³

¹Office of Surveillance and Epidemiology/DDRE, Silver Spring, Maryland, USA; ²Office of Surveillance and Epidemiology/DSRCS, Silver Spring, Maryland, USA; ³Office of Surveillance and Epidemiology/SARA, Rockville, Maryland, USA.

INTRODUCTION

The United States Food and Drug Administration (FDA) is the Federal public health agency that has regulatory responsibility for ensuring the safety of all marketed medical products, including pharmaceuticals (i.e., drugs and biologics) (see also Chapter 6). In order to ensure that safe and effective pharmaceuticals are available, the FDA relies on both the recognition, and *voluntary* reporting, of serious adverse events (AEs) by health care providers and their patients and the *mandatory* reporting of AEs by manufacturers as required by law and regulation.

All unsolicited reports from health care professionals or consumers, received by the FDA via either the voluntary or mandatory route, are called *spontaneous* reports. A spontaneous report is a clinical observation that originates outside of a formal study. The individual spontaneous reports of adverse drug reactions (ADRs), medication errors, and product quality problems, sent directly to the FDA through the MedWatch program (see below) or to the manufacturer and then indirectly from the manufacturer to the FDA, combined with data from formal clinical studies and from the medical

and scientific literature, comprise the primary data source upon which postmarketing surveillance depends. In the US, a large majority of reports, between 70% and 75%, are submitted either directly or indirectly by health care professionals as voluntary reports, with consumer/patient reports comprising about 15% of reports.

In addition to this passive process for safety surveillance, the FDA continues to explore the use of new active surveillance methodologies for collecting reports of adverse effects and evaluating adverse events. The FDA may also explore drug safety questions in large population-based claim databases that link prescriptions with adverse outcomes.

When the FDA approves a pharmaceutical product for prescribing and dispensing by health care providers in the United States, the agency has conducted a rigorous, science-based, multidisciplinary review of controlled clinical trials sponsored and conducted by a pharmaceutical company. The FDA has determined that the product's benefits outweigh any known or anticipated risks for the general population when the product is used as indicated in the approved prescribing information. However, the limitations inherent in the controlled clinical trial setting in

the identification of rare, but clinically important, adverse events inevitably insure that uncertainties will remain about the safety of the pharmaceutical once it is marketed and used in a wider population, over longer periods of time, in patients with comorbidities and concomitant medications, and for “off-label” uses not previously evaluated.

Given these recognized and accepted limitations in the pre-approval New Drug Application (NDA) process, the agency relies on the public, both health care professionals and their patients, for the voluntary reporting of suspected, serious, and unlabeled ADRs, medication errors, and product quality problems observed during the use of the pharmaceutical in the “real-world” setting, in order to manage the risk of product use and reduce the possibility of harm to patients.

Harm to patients from pharmaceutical use may occur due to four types of risk (Figure 7.1). Most injuries and deaths associated with the use of medical products result from their *known side effects*, some unavoidable but others able to be prevented or minimized by careful product choice and use. It is estimated that more than half the side effects of pharmaceuticals are avoidable.

Other sources of preventable adverse events are *medication errors*, which may occur when the product is administered incorrectly or when the wrong drug or dose is administered.

Injury from *product quality problems* is of interest to the FDA, which has regulatory responsibility for oversight of product quality control and quality assurance during the manufacturing and distribution process.

The final category of potential risk, those risks most amenable to identification by an effective voluntary reporting system, involves the *remaining uncertainties* about a product. These uncertainties include unexpected and rare AEs, long-term effects, unstudied uses and/or unstudied

populations, unanticipated medication errors due to name confusion or packaging format, and product quality defects during the manufacturing process.

This chapter reviews the history of AE reporting in the United States, its terminology, and its regulatory aspects. The strengths, limitations, and applications of the FDA’s Adverse Event Reporting System (AERS) are discussed, as are future plans.

HISTORY OF US PHARMACEUTICAL SAFETY REGULATION

The FDA is the first US consumer protection agency. Its predecessor, the Bureau of Drugs, was established in order to implement the Biologics Control Act of 1902. Subsequent drug regulatory laws, in 1906, 1938, and 1962, have all resulted from widespread public concern about drug safety and demands that the US Congress address a perceived crisis that threatened the health and lives of children. Each law or amendment incrementally strengthened the FDA’s capability to effectively monitor the postmarketing safety of drugs and other medical products.

The 1902 Act was passed by the US Congress in reaction to the public outrage from hundreds of cases of post-vaccination tetanus and the deaths of several dozen children due to tetanus-contaminated diphtheria antitoxin. This first drug safety law required annual licensing of manufacturers and distributors and the labeling of all products with the name of the manufacturer. Neither the premarketing safety and efficacy nor the postmarketing safety of these products was regulated by the government.

The Pure Food and Drug Act of 1906 prohibited interstate commerce of mislabeled and adulterated drugs and

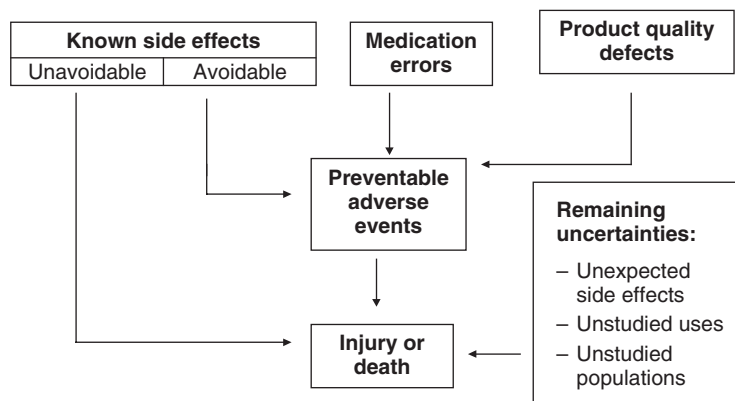


Figure 7.1. Sources of risk from medical products.

foods. Again, the safety of drugs after consumption was not addressed. For example, in 1934 the Agency began investigations on products containing dinitrophenol, a component in diet preparations that increased metabolic rate to dangerous levels, and was responsible for many deaths and injuries. However, the Office of Drug Control could not seize the products, and was limited to posting warnings. The safety of drugs after consumption was not addressed until the 1930s and unfortunately it was again a disaster that prompted Congress to act.

These continuing problems with dangerous drugs that fell outside the controls of the Pure Food and Drug Act finally received national attention with the elixir of sulfanilamide disaster in 1937. The S.E. Massengill Co. introduced a flavorful oral dosage form of the new anti-infective “wonder drug” by using an untested solvent, the antifreeze diethylene glycol. By the time the FDA became aware of the problem and removed the product from pharmacy shelves and medicine cabinets, the preparation had caused 107 deaths, including many children. Even though the toxic effects of diethylene glycol were well documented by 1931, with no drug safety regulations in place the only charge that could be brought under the 1906 Act was misbranding the product, since there was no alcohol in the “elixir,” as implied by the name.

In June 1938, the Federal Food, Drug and Cosmetic Act was passed by the US Congress. The law required new drugs to be tested for safety before marketing, the results of which would be submitted to the FDA in an NDA. The law also required that drugs have adequate labeling for safe use. Again, no postmarketing safety monitoring was mandated by this new law.

During the 1950s, there was a rapid expansion of the pharmaceutical industry and an increase in the number of new products. A new broad spectrum antibiotic, chloramphenicol, was approved by the FDA in early 1949 as “safe and effective when used as indicated” (the standard for approval in the 1938 Act). However, the small number of patients exposed to chloramphenicol during pre-approval clinical trials was not adequate to observe serious but rare adverse events that would occur in fewer than 1 in 1000 patients. Within six months of approval, reports in the medical literature in the US and Europe suggested the association of fatal aplastic anemia with chloramphenicol use.

In late June 1952, in order to gather the necessary data to evaluate this issue, the FDA ordered the staff in all 16 district offices to contact every hospital, medical school, and clinic in cities with populations of 100 000 or more to collect information on any cases of aplastic anemia or other blood dyscrasias attributed to chloramphenicol. Within

four days of field contacts, an additional 217 cases of chloramphenicol-associated blood dyscrasias had been identified.

The delay in identification of and regulatory action on reports of aplastic anemia associated with chloramphenicol use demonstrated the necessity for monitoring adverse events following the approval and marketing of new drugs. In response to this need the American Medical Association (AMA) established a Committee on Blood Dyscrasias, which began collecting case reports of drug-induced blood-related illness in 1954. At that time, the AMA had a potential information source of over 7000 hospitals and 250 000 physicians. The AMA’s program was expanded in 1961 to a more comprehensive “Registry on Adverse Reactions.” The program was discontinued in 1971 because of parallel efforts by the FDA.

In 1956, the FDA piloted its own drug ADR surveillance program in cooperation with the American Society of Hospital Pharmacists (the predecessor of the American Society of Health-System Pharmacists), the national association of medical records librarians, and the AMA. The reporting program began with 6 hospitals and by 1965 had grown to over 200 teaching hospitals which reported to the FDA on a monthly basis. In addition, reports were sent to the FDA from selected Federal hospitals (Department of Defense, Veterans Administration, Public Health Service) and published reports were culled from the medical literature and received from the World Health Organization.

The 1962 Kefauver–Harris Amendments to the Food, Drug, and Cosmetic Act of 1938 required proof of efficacy before drug approval and marketing. For the first time, this law also mandated that pharmaceutical manufacturers must report AEs to the FDA for any of their products having an NDA—the vast majority of prescription products introduced since 1938.

The FDA began to computerize the storage of its AE reports in 1967, and by early 1968 received, coded, and entered all data from the FDA Form 1639 Drug Experience Reports into the Spontaneous Reporting System (SRS). The SRS was replaced in November 1997 with the Adverse Event Reporting System (AERS), a computerized information database that supports the FDA’s postmarketing safety surveillance program for all approved drug and therapeutic biologic products. The AERS is an internationally compatible system designed as a pharmacovigilance tool for storing and analyzing safety reports.

By 1991, there were five different forms for manufacturers and health professionals to report medical product problems to the agency. In 1993, then-FDA Commissioner David A. Kessler, MD, citing confusion with the multiple forms,

launched the FDA's MedWatch Adverse Event Reporting Program. A single-page voluntary reporting form, FDA form 3500 (The "MedWatch" form), was introduced to report adverse events associated with all medical products except vaccines, and the FDA form 3500A was provided for use by mandatory reporters (see Figures 7.2 and 7.3). The MedWatch program was charged with the task of facilitating, supporting, and promoting the voluntary reporting process. Since 1993, over 200 000 voluntary reports have been received from health care professionals and consumers, coded, and entered into the FDA AERS database (see Figure 7.4).

REGULATORY REPORTING REQUIREMENTS

In the US, AE reporting by individual health care providers and consumers is voluntary. However, manufacturers, packers, and distributors of FDA-approved pharmaceuticals (drugs and biologic products) all have mandatory reporting requirements governed by regulation. Historically, only nonbiologic pharmaceutical products with approved NDAs (i.e., all prescription and some over-the-counter drugs) were subject to mandatory reporting requirements. In 1994, this requirement was expanded to include biologic products.

It should be emphasized that these regulations are aimed at pharmaceutical manufacturers, but also provide a useful framework for reporting by practitioners to either the FDA and/or the manufacturer. In the US, most health professionals and consumers report AEs to the manufacturer rather than directly to the FDA. This pattern is not seen in many other countries, where consumers and health professionals report directly to a governmental public health agency.

CURRENT REQUIREMENTS

The main objective of the FDA postmarketing reporting requirement is to provide early prompt detection of signals about potentially serious, previously unknown safety problems with marketed drugs, especially with newly marketed drugs. To understand the regulatory requirements, one first needs to define several terms. These definitions are revisions that became effective in April 1998.

An *adverse experience* is any AE associated with the use of a drug or biologic product in humans, whether or not considered product related, including the following: an AE occurring in the course of the use of the product in professional practice, an AE occurring from overdose of the product, whether accidental or intentional, an AE occurring from abuse of the product, an AE occurring from withdrawal

of the product, and any failure of expected pharmacologic action.

An *unexpected adverse experience* means any AE that is not listed in the current labeling for the product. This includes events that may be symptomatically and pathophysiologically related to an event listed in the labeling, but differ from the event because of greater severity or specificity.

A *serious adverse experience* is any AE occurring at any dose that results in any of the following outcomes: death, a life-threatening AE, inpatient hospitalization or prolongation of existing hospitalization, a persistent or significant disability/incapacity, or congenital anomaly/birth defect. Important medical events that may not result in death, may not be life-threatening, or may not require hospitalization may be considered a serious AE when, based upon appropriate medical judgment, they may jeopardize the patient or subject and may require medical or surgical intervention to prevent one of the outcomes listed in this definition. Examples of such medical events include allergic bronchospasm requiring intensive treatment in an emergency room or at home, blood dyscrasias or convulsions that do not result in inpatient hospitalization, or the development of drug dependency or drug abuse.

Table 7.1 outlines the US mandatory reporting requirements regarding pharmaceuticals. By regulation, companies are required to report to the FDA *all* adverse events of which they become aware and to provide as complete information as possible. Although pharmaceutical reporting is mandated, it still relies primarily on information provided to them by health professionals through both voluntary reporting and the scientific literature.

In the case of over-the-counter (OTC) drugs, reports are only required on OTC products marketed under an approved NDA, including those prescription drugs that undergo a switch to OTC status. Reports are not currently required for other OTC drugs (i.e., older drug ingredients which are marketed without an NDA), although voluntary reporting is encouraged for serious events.

Both prescription and OTC drugs require FDA safety and efficacy review prior to marketing, unlike dietary supplements (which include vitamins, minerals, amino acids, botanicals, and other substances used to increase total dietary intake). By law, the manufacturers of these latter products do not have to prove safety or efficacy, but that same law places the responsibility on the FDA to demonstrate that a particular product is unsafe or presents a potentially serious risk to public health. In addition, manufacturers of these products do not have to report AEs to the FDA. As a result, direct-to-FDA voluntary reporting by health professionals and their patients of

U.S. Department of Health and Human Services

Form Approved: OMB No. 0910-0291, Expires: 03/31/05
See OMB statement on reverse.**MEDWATCH**For VOLUNTARY reporting of
adverse events and product problemsThe FDA Safety Information and
Adverse Event Reporting Program

Page ____ of ____

FDA USE ONLYTriage unit
sequence #**A. PATIENT INFORMATION**

1. Patient Identifier	2. Age at Time of Event: or _____ Date of Birth: _____	3. Sex ☐ Female ☐ Male	4. Weight ____ lbs or ____ kgs

B. ADVERSE EVENT OR PRODUCT PROBLEM

1. ☐ Adverse Event and/or ☐ Product Problem (e.g., defects/malfunctions)	
2. Outcomes Attributed to Adverse Event (Check all that apply) ☐ Death: _____ (mo/day/yr) ☐ Life-threatening ☐ Hospitalization - initial or prolonged	
☐ Disability ☐ Congenital Anomaly ☐ Required Intervention to Prevent Permanent Impairment/Damage ☐ Other: _____	
3. Date of Event (mo/day/year)	4. Date of This Report (mo/day/year)

5. Describe Event or Problem

6. Relevant Tests/Laboratory Data, Including Dates

7. Other Relevant History, including Preexisting Medical Conditions (e.g., allergies, race, pregnancy, smoking and alcohol use, hepatic/renal dysfunction, etc.)

C. SUSPECT MEDICATION(S)

1. Name (Give labeled strength & mfr/labeler, if known) #1 _____ #2 _____	
2. Dose, Frequency & Route Used #1 _____ #2 _____	3. Therapy Dates (If unknown, give duration) from/to (or best estimate) #1 _____ #2 _____
4. Diagnosis for Use (Indication) #1 _____ #2 _____	5. Event Abated After Use Stopped or Dose Reduced? #1 ☐ Yes ☐ No ☐ Doesn't Apply #2 ☐ Yes ☐ No ☐ Doesn't Apply
6. Lot # (if known) #1 _____ #2 _____	7. Exp. Date (if known) #1 _____ #2 _____
8. Event Reappeared After Reintroduction? #1 ☐ Yes ☐ No ☐ Doesn't Apply #2 ☐ Yes ☐ No ☐ Doesn't Apply	
9. NDC# (For product problems only) #1 _____ #2 _____	
10. Concomitant Medical Products and Therapy Dates (Exclude treatment of event)	

D. SUSPECT MEDICAL DEVICE

1. Brand Name		
2. Type of Device		
3. Manufacturer Name, City and State		
4. Model #	Lot #	5. Operator of Device ☐ Health Professional ☐ Lay User/Patient ☐ Other: _____
Catalog #	Expiration Date (mo/day/yr)	
Serial #	Other #	
6. If Implanted, Give Date (mo/day/yr)		7. If Explanted, Give Date (mo/day/yr)
8. Is this a Single-use Device that was Reprocessed and Reused on a Patient? ☐ Yes ☐ No		
9. If Yes to Item No. 8, Enter Name and Address of Reprocessor		
10. Device Available for Evaluation? (Do not send to FDA) ☐ Yes ☐ No ☐ Returned to Manufacturer on: _____ (mo/day/yr)		
11. Concomitant Medical Products and Therapy Dates (Exclude treatment of event)		

E. REPORTER (See confidentiality section on back)

1. Name and Address		Phone #
2. Health Professional? ☐ Yes ☐ No		
3. Occupation		4. Also Reported to: ☐ Manufacturer ☐ User Facility ☐ Distributor/Importer
5. If you do NOT want your identity disclosed to the manufacturer, place an "X" in this box: ☐		

Mail to: **MEDWATCH**
5600 Fishers Lane
Rockville, MD 20852-9787-or- FAX to:
1-800-FDA-0178

FORM FDA 3500 (12/03) Submission of a report does not constitute an admission that medical personnel or the product caused or contributed to the event.

Figure 7.2. MedWatch Voluntary Reporting Form (FDA Form 3500).

ADVICE ABOUT VOLUNTARY REPORTING

Report adverse experiences with:

- Medications (*drugs or biologics*)
- Medical devices (*including in-vitro diagnostics*)
- Special nutritional products (*dietary supplements, medical foods, infant formulas*)
- Cosmetics
- Medication errors

Report product problems — quality, performance or safety concerns such as:

- Suspected counterfeit product
- Suspected contamination
- Questionable stability
- Defective components
- Poor packaging or labeling
- Therapeutic failures

Report **SERIOUS** adverse events. An event is serious when the patient outcome is:

- Death
- Life-threatening (*real risk of dying*)
- Hospitalization (*initial or prolonged*)
- Disability (*significant, persistent or permanent*)
- Congenital anomaly
- Required intervention to prevent permanent impairment or damage

Report even if:

- You're not certain the product caused the event
- You don't have all the details

How to report:

- Just fill in the sections that apply to your report
- Use section C for all products except medical devices
- Attach additional blank pages if needed
- Use a separate form for each patient
- Report either to FDA or the manufacturer (*or both*)

Confidentiality: The patient's identity is held in strict confidence by FDA and protected to the fullest extent of the law. FDA will not disclose the reporter's identity in response to a request from the public, pursuant to the Freedom of Information Act. The reporter's identity, including the identity of a self-reporter, may be shared with the manufacturer unless requested otherwise.

If your report involves a serious adverse event with a device and it occurred in a facility outside a doctor's office, that facility may be legally required to report to FDA and/or the manufacturer. Please notify the person in that facility who would handle such reporting.

Important numbers:

- 1-800-FDA-0178 -- To FAX report
- 1-800-FDA-1088 -- To report by phone or for more information
- 1-800-822-7967 -- For a VAERS form for vaccines

To Report via the Internet:

<http://www.fda.gov/medwatch/report.htm>

The public reporting burden for this collection of information has been estimated to average 30 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to:

*Department of Health and Human Services
Food and Drug Administration
MedWatch; HFD-410
5600 Fishers Lane
Rockville, MD 20857*

*Please DO NOT
RETURN this form
to this address.*

OMB statement:

"An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number."

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Food and Drug Administration

FORM FDA 3500 (12/03) (Back)

Please Use Address Provided Below -- Fold in Thirds, Tape and Mail

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Food and Drug Administration
Rockville, MD 20857

Official Business
Penalty for Private Use \$300



NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES
OR APO/FPO

BUSINESS REPLY MAIL

FIRST CLASS MAIL PERMIT NO. 946 ROCKVILLE MD

POSTAGE WILL BE PAID BY FOOD AND DRUG ADMINISTRATION

MEDWATCH

The FDA Safety Information and Adverse Event Reporting Program
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20852-9787



Figure 7.2. (Continued).

U.S. Department of Health and Human Services

MEDWATCHThe FDA Safety Information and
Adverse Event Reporting ProgramFor use by user-facilities,
importers, distributors and manufacturers
for MANDATORY reporting

Page ____ of ____

Form Approved: OMB No. 0910-0291, Expires: 03/31/05
See OMB statement on reverse

Mfr Report #

UF/Importer Report #

FDA Use Only

A. PATIENT INFORMATION				C. SUSPECT MEDICATION(S)			
1. Patient Identifier	2. Age at Time of Event: or Date of Birth:	3. Sex ☐ Female ☐ Male	4. Weight ____ lbs or ____ kgs	1. Name (Give labeled strength & mfr/labeler, if known) #1 _____ #2 _____			
In confidence				2. Dose, Frequency & Route Used #1 _____ #2 _____		3. Therapy Dates (If unknown, give duration) from/to (or best estimate) #1 _____ #2 _____	
B. ADVERSE EVENT OR PRODUCT PROBLEM				D. SUSPECT MEDICAL DEVICE			
1. ☐ Adverse Event and/or ☐ Product Problem (e.g., defects/malfunctions)				1. Brand Name _____			
2. Outcomes Attributed to Adverse Event (Check all that apply) ☐ Death: _____ (mo/day/yr) ☐ Life-threatening ☐ Hospitalization - initial or prolonged				2. Type of Device _____			
☐ Disability ☐ Congenital Anomaly ☐ Required Intervention to Prevent Permanent Impairment/Damage ☐ Other: _____				3. Manufacturer Name, City and State _____			
3. Date of Event (mo/day/year)				4. Model # _____ Lot # _____			
4. Date of This Report (mo/day/year)				5. Operator of Device ☐ Health Professional ☐ Lay User/Patient ☐ Other: _____			
5. Describe Event or Problem				6. Catalog # _____ Expiration Date (mo/day/yr) _____			
6. Relevant Tests/Laboratory Data, Including Dates				7. Serial # _____ Other # _____			
7. Other Relevant History, Including Preexisting Medical Conditions (e.g., allergies, race, pregnancy, smoking and alcohol use, hepatic/renal dysfunction, etc.)				8. If Implanted, Give Date (mo/day/yr) _____ 9. If Explanted, Give Date (mo/day/yr) _____			
				10. Is this a Single-use Device that was Reprocessed and Reused on a Patient? ☐ Yes ☐ No			
				11. Device Available for Evaluation? (Do not send to FDA) ☐ Yes ☐ No ☐ Returned to Manufacturer on: _____ (mo/day/yr)			
				12. Concomitant Medical Products and Therapy Dates (Exclude treatment of event)			
E. INITIAL REPORTER							
1. Name and Address						Phone # _____	
2. Health Professional? ☐ Yes ☐ No						3. Occupation _____	
4. Initial Reporter Also Sent Report to FDA ☐ Yes ☐ No ☐ Unk.							



Submission of a report does not constitute an admission that medical personnel, user facility, importer, distributor, manufacturer or product caused or contributed to the event.

FORM FDA 3500A (9/03)

Figure 7.3. MedWatch Mandatory Reporting Form (FDA Form 3500A).

Medication and Device Experience Report

(Continued)

Submission of a report does not constitute an admission that medical personnel, user facility, importer, distributor, manufacturer or product caused or contributed to the event.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service • Food and Drug Administration

Refer to guidelines for specific instructions.

Page ____ of ____

F. FOR USE BY USER FACILITY/IMPORTER (Devices Only)			
1. Check One ☐ User Facility ☐ Importer		2. UF/Importer Report Number	
3. User Facility or Importer Name/Address			
4. Contact Person		5. Phone Number	
6. Date User Facility or Importer Became Aware of Event (mo/day/yr)		8. Date of This Report (mo/day/yr)	
		7. Type of Report ☐ Initial ☐ Follow-up # _____	
9. Approximate Age of Device		10. Event Problem Codes (Refer to coding manual)	
		Patient Code _____ - _____ - _____ Device Code _____ - _____ - _____	
11. Report Sent to FDA? ☐ Yes _____ (mo/day/yr) ☐ No		12. Location Where Event Occurred ☐ Hospital ☐ Outpatient Diagnostic Facility ☐ Home ☐ Ambulatory Surgical Facility ☐ Nursing Home ☐ Outpatient Treatment Facility ☐ Other: _____ (Specify)	
13. Report Sent to Manufacturer? ☐ Yes _____ (mo/day/yr) ☐ No			
14. Manufacturer Name/Address			
G. ALL MANUFACTURERS			
1. Contact Office - Name/Address (and Manufacturing Site for Devices)		2. Phone Number	
		3. Report Source (Check all that apply) ☐ Foreign ☐ Study ☐ Literature ☐ Consumer ☐ Health Professional ☐ User Facility ☐ Company Representative ☐ Distributor ☐ Other: _____	
4. Date Received by Manufacturer (mo/day/yr)		5. (A)NDA # _____ IND # _____ PLA # _____ Pre-1938 ☐ Yes OTC Product ☐ Yes	
6. If IND, Give Protocol #			
7. Type of Report (Check all that apply) ☐ 5-day ☐ 15-day ☐ 10-day ☐ Periodic ☐ Initial ☐ Follow-up # _____		8. Adverse Event Term(s)	
9. Manufacturer Report Number			

The public reporting burden for this collection of information has been estimated to average one hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to:

FORM FDA 3500A (9/03) (Back)

H. DEVICE MANUFACTURERS ONLY	
1. Type of Reportable Event ☐ Death ☐ Serious Injury ☐ Malfunction ☐ Other: _____	
2. If Follow-up, What Type? ☐ Correction ☐ Additional Information ☐ Response to FDA Request ☐ Device Evaluation	
3. Device Evaluated by Manufacturer? ☐ Not Returned to Manufacturer ☐ Yes ☐ Evaluation Summary Attached ☐ No (Attach page to explain why not) or provide code: _____	
4. Device Manufacture Date (mo/yr)	
5. Labeled for Single Use? ☐ Yes ☐ No	
6. Evaluation Codes (Refer to coding manual)	
Method _____ - _____ - _____ - _____	
Results _____ - _____ - _____ - _____	
Conclusions _____ - _____ - _____ - _____	
7. If Remedial Action Initiated, Check Type ☐ Recall ☐ Notification ☐ Repair ☐ Inspection ☐ Replace ☐ Patient Monitoring ☐ Relabeling ☐ Modification/Adjustment ☐ Other: _____	
8. Usage of Device ☐ Initial Use of Device ☐ Reuse ☐ Unknown	
9. If action reported to FDA under 21 USC 360(f), list correction/removal reporting number: _____	
10. ☐ Additional Manufacturer Narrative and / or 11. ☐ Corrected Data	

Department of Health and Human Services
Food and Drug Administration
MedWatch; HFD-410
5600 Fishers Lane
Rockville, MD 20857

Please DO NOT RETURN this form to this address.

OMB Statement:
"An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number."

Figure 7.3. (Continued).

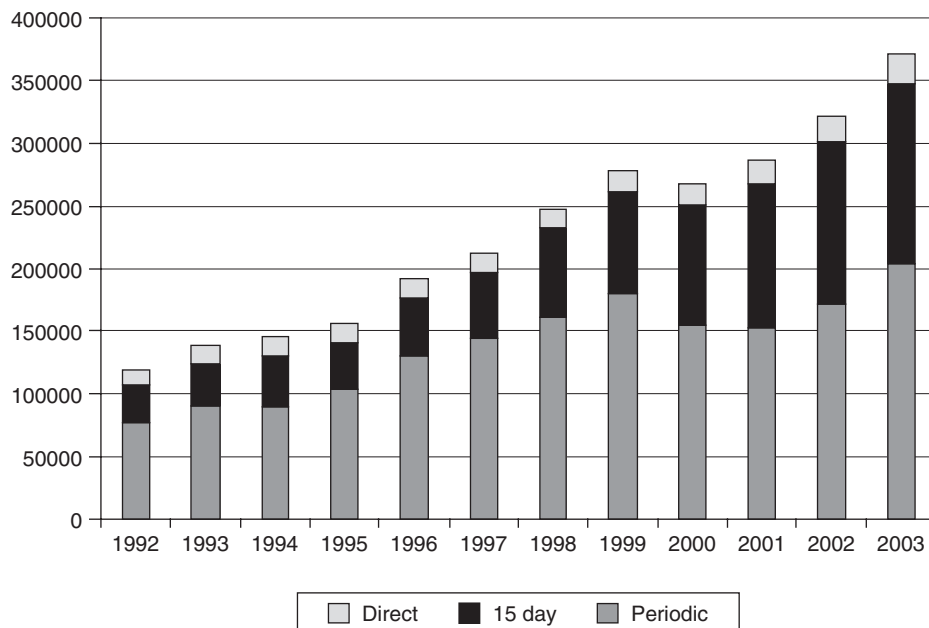


Figure 7.4. Adverse drug event reports by year and type 1992–2003.

Table 7.1. Mandatory adverse event (AE) reporting requirements for pharmaceuticals

15-day "Alert Reports"	All serious and unexpected AEs, whether foreign or domestic, must be reported to the FDA within 15 calendar days
15-day "Alert Reports" follow-up	The manufacturer must promptly investigate all AEs that are the subject of a 15-day Alert Report and submit a follow-up report within 15 calendar days
Periodic AE reports	All non-15-day, domestic AE reports must be reported periodically (quarterly for the first 3 years after approval, then annually). Periodic reports for products marketed prior to 1938 are not required. Periodic reporting does not apply to AE information obtained from postmarketing studies or from reports in the scientific literature
Scientific literature	A 15-day Alert Report based on information from the scientific literature (case reports or results from a formal clinical trial). A copy of the published article must accompany the report, translated into English if foreign
Postmarketing studies	No requirement for a 15-day Alert Report on an AE acquired from a postmarketing study <i>unless</i> the manufacturer concludes a reasonable possibility that the product caused the event

serious adverse events associated with and possibly causally linked to dietary supplements is particularly important. To help promote reporting and tracking of adverse events associated with dietary supplements, the FDA's Center for Food Safety and Nutrition (CFSAN) launched its CFSAN Adverse Event Reporting System (CAERS) in the summer of 2003.

The specific regulations governing postmarketing AE reporting by pharmaceutical companies are listed in Table 7.2. Accompanying separate guidances for drugs and biologics

were made available in 1992 and 1993, respectively. As can be seen, the regulations have each been amended numerous times.

Many of the proposed rules, draft guidance documents, and a docket memo (in various stages of development) encourage electronic AE reporting. Electronic reporting is an important step because reports are available for review more quickly. Further, electronic reporting reduces data entry costs, allowing the Center for Drug Evaluation and

Table 7.2. Federal regulations regarding postmarketing adverse event reporting

21 CFR 310.305 Prescription drugs not subject to premarket approval	[July 3, 1986 (51 FR 24779), amended October 13, 1987 (52 FR 37936); March 29, 1990 (55 FR 1578); April 28, 1992 (57 FR 17980); June 25, 1997 (62 FR 34167); October 7, 1997 (62 FR 52249); March 4, 2002 (67 FR 9585)]
21 CFR 314.80 Human drugs with approved new drug applications (NDAs)	[February 22, 1985 (50 FR 7493) and April 11, 1985 (50 FR 14212), amended May 23, 1985 (50 FR 21238); July 3, 1986 (51 FR 24481); October 13, 1987 (52 FR 37936); March 29, 1990 (55 FR 11580); April 28, 1992 (57 FR 17983); June 25, 1997 (62 FR 34166, 34168); October 7, 1997 (62 FR 52251); March 26, 1998 (63 FR 14611); March 4, 2002 (67 FR 9586)]
21 CFR 314.98 Human drugs with approved abbreviated new drug applications (ANDAs)	[April 28, 1992 (57 FR 17983), amended January 5, 1999 (64 FR 401)]
21 CFR 600.80 Biological products with approved product license applications (PLAs)	[October 27, 1994 (59 FR 54042), amended June 25, 1997 (62 FR 34168); October 7, 1997 (62 FR 52252); March 26, 1998 (63 FR 14612); October 20, 1999 (64 FR 56449)]

Research (CDER) to use its resources for additional pharmacovigilance efforts. The proposed rules, draft guidances, and docket memo and their associated statutes are as follows:

- The proposed rule on Adverse Event Reporting and guidance on electronic submissions are currently being finalized.
- A draft Guidance for Industry, “Providing Regulatory Submissions in Electronic Format—Postmarketing Expedited Safety Reports,” was released in May 2001.
- A memo entitled “Postmarketing Expedited Safety Reports—15-Day Alert Reports” was added to public Docket 92S-0251 on May 22, 2002. This memo allows for voluntary electronic reporting of 15-day (expedited) safety reports with no paper submissions required.
- A draft Guidance for Industry entitled “Providing Regulatory Submissions in Electronic Format—Postmarketing Periodic Adverse Drug Experience Reports” was published on June 24, 2003.

As of the end of 2003, nearly 20% of all expedited reports were submitted electronically, and the FDA encourages firms to participate in this voluntary process, replacing MedWatch (3500A) reports. To facilitate this effort, the FDA hosts a meeting twice yearly with representatives from major pharmaceutical firms. The purpose of this meeting is to discuss electronic AE reporting, including ways to stimulate increased electronic reporting within the industry. A description of the process of how the FDA handles these reports will be provided in a later section of this chapter.

Recent Changes

In recent years, there has been a significant international effort to standardize the pharmaceutical regulatory environment worldwide through the auspices of the International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use. These efforts toward international harmonization have a direct impact on how the FDA is currently rewriting regulations on AE reporting. The AERS was launched in November 1997 and is an internationally compatible system in full accordance with the ICH initiatives.

The initiatives that directly affect postmarketing surveillance are:

- M1 IMT (International Medical Terminology): the AERS uses the Medical Dictionary for Regulatory Activities (MedDRA) as its coding tool for reported adverse reaction/adverse event terms via individual case safety reports.
- M2 ESTR (Electronic Standards for the Transfer of Regulatory Information): the AERS uses ESTR standards for submission of individual case safety reports in electronic form via the electronic data interchange (EDI) gateway.
- E2B(M) (Data Elements for Transmission of Individual Case Safety Reports): the AERS has implemented the E2B data format into its database, and will use the E2B as the standard for electronic submissions.
- E2C PSUR (Periodic Safety Update Reports): defines a standard format for clinical safety data management for PSURs for marketed drugs. Initially, PSURs will be submitted on paper, and the FDA has published guidance to allow these summaries to be sent in electronically to the electronic Central Document Room (eCDR).

The Agency has undertaken a major effort in implementation of the electronic reporting of Individual Case Safety Reports (ICSRs) based on the ICH E2B(M), M1 (MedDRA), and M2 standards, and to clarify and revise its regulations regarding pre- and post-marketing safety reporting requirements for human drug and biologic products. In the *Federal Register* of October 7, 1997 (62 FR 52237), the FDA published a final rule amending its regulations for expedited safety reporting. This final rule implements the ICH E2A initiative on clinical safety data management. Based on E2A, the final rule provides an internationally accepted definition of “serious,” requires the submission of the MedWatch 3500A for paper submissions, requires expedited reports in a 15 calendar rather than working day time frame, and harmonizes procedures for reporting pre- and post-marketing as well as international and domestic reporting. With regard to the postmarketing safety reporting regulations for human drug and licensed biologic products, the Agency published a proposed rule in the *Federal Register* of October 27, 1994 (59 FR 54046), to amend these requirements (as well as others), to implement international standards, and to facilitate the reporting of adverse experiences.

To help the pharmaceutical manufacturers understand the new requirements, on August 27, 1998 the FDA published an interim guidance for industry, “Postmarketing Adverse Experience Reporting for Human Drugs and Licensed Biologic Products: Clarification of What to Report.”

In the *Federal Register* of November 5, 1998 (63 FR 59746), the Agency published an Advanced Notice of Proposed Rulemaking to notify manufacturers that it is considering preparing a proposed rule that would require them to submit individual case reports electronically using standardized medical terminology, standardized data elements, and electronic transmission standards as recommended by the ICH in the M1, M2, and E2B(M) initiatives. The FDA published a Public Docket 92A-0251, “Electronic Submission of Postmarketing Expedited Periodic Individual Case Safety Reports,” which allows pharmaceutical companies to submit reports to the FDA electronically.

In March 2001, the Agency issued a “Guidance for Industry: Postmarketing Safety Reporting for Human Drug and Biological Products Including Vaccines,” which superseded the March 1992 document.

The November 2001 Guidance for Industry “Electronic Submission of Postmarketing Expedited Safety Reports” describes how pharmaceutical companies may submit ICSRs using EDI gateway and physical media (e.g., CD-ROM) and attachments to ICSRs using only physical media.

In May 2002, the FDA issued a Guidance for Industry “Providing Regulatory Submissions in Electronic Format—Postmarketing Periodic Adverse Drug Experience Reports,” which describes how pharmaceutical companies may submit periodic ICSRs with and without attachments and descriptive information (including PSURs) using physical media.

In September 2003, the FDA issued a Guidance for Industry “Providing Regulatory Submissions in Electronic Format—Annual Reports for NDAs and ANDAs,” which describes how pharmaceutical companies may submit descriptive information (including PSURs) using physical media.

On October 1, 2003, the FDA transferred certain product oversight responsibilities from the Center for Biologics Evaluation and Research (CBER) to the CDER. This consolidation provides greater opportunities to further develop and coordinate scientific and regulatory activities between the CBER and the CDER, leading to a more efficient and consistent review program for human drugs and biologics. The FDA believes that as more drugs and biologic products are developed for a broader range of illnesses, such interaction is necessary for both efficient and consistent agency action. Under the new structure, the biologic products transferred to the CDER will continue to be regulated as licensed biologics.

PROPOSED MODIFICATIONS

At the current time, the FDA is working on further modifications to the postmarketing safety reporting requirements. A Serious Adverse Drug Reaction (SADR) Reporting Proposed Rule is expected to be published in the near future, that focuses on report quality, standardizes terminology to “Adverse Drug Reaction,” and encourages active query by a health care professional at the company who speaks directly with the initial reporter of the serious adverse reaction report. This entails, at a minimum, a focused line of questioning designed to capture clinically relevant information, follow-up, and determination of seriousness, and defines the minimum data set for safety reports. The proposed rule will also implement the ICH E2C: International PSUR, which contains marketing status, core labeling (company core data sheet (CCDS); company core safety information (CCSI) is safety information in CCDS), changes in safety status since last report, exposure data, clinical explanation of cases, data line (narrative summary of the individual case safety reports which provide demographic, drug, and event information) listings and tables, status of postmarketing surveillance safety studies, overall critical analyses, and assessments. The earlier October 27, 1994

proposed amendments to the postmarketing periodic AE reporting requirements will be repropounded in this current Proposed Rule, based on a guidance on this topic developed by the ICH.

As noted previously, OTC products without an NDA are not subject to reporting. To bring these products into the postmarketing safety net, the FDA plans to publish an OTC ADR Reporting Proposed Rule. Consideration is being given to the requirement for ADR reporting for OTC monograph drugs, since most marketed OTC drugs lack an approved NDA. The FDA's review of marketed OTC drugs without approved NDAs (ANDAs) has been accomplished through rulemaking, establishing conditions in OTC drug monographs for drugs within therapeutic classes (e.g., laxatives). An OTC drug monograph specifies the conditions (i.e., ingredients and concentrations, testing procedures, dosage, labeling, and mode of administration) under which an OTC drug is generally recognized as safe and effective and is not misbranded.

In an effort to expand the Agency's ability to monitor and improve the safe use of human and biologic products both during clinical trials and once the products are on the market, the FDA on March 14, 2003 published a proposed rule, titled "Safety Reporting Requirements for Human Drug and Biological Products" (*The Tome*), which would require companies to file expedited reports of suspected ADRs unless the company is certain the product did not cause the reaction.

The Tome recommended the replacement of periodic drug adverse experience reports (21 CFR 314.80) with PSURs. Currently, the CDER encourages industry to submit a waiver to allow submission of PSURs instead of periodic drug adverse experience reports. PSURs are in the format proposed by the ICH of Technical Requirements for Registration of Pharmaceuticals for Human Use, Topic E2C. The PSUR summarizes the safety data received by a sponsor for an application from worldwide sources for a specific time frame. The number of PSURs received is dependent on the number of NDAs/ANDAs marketed. The PSUR format enhances postmarketing drug and therapeutic biologic safety because it requires additional information and analyses (such as patient exposure data) not required in the periodic adverse drug experience report. These additional data enhance our review of postmarketing safety.

DATA COLLECTION: THE MEDWATCH PROGRAM

An effective national postmarketing surveillance system depends on voluntary reporting of adverse events,

medication errors, and product quality problems by health professionals and consumers to the FDA, either directly or via the manufacturer. Neither individual health professionals nor hospitals are required by Federal law or regulation to submit AE reports on pharmaceuticals, although Federal law does require hospitals and other "user facilities" to report deaths and serious injuries that occur with medical devices.

Many health care organizations recommend and promote the reporting of AEs to the FDA. Adverse event monitoring by hospitals is included in the Joint Commission on the Accreditation of Health Care Organizations (JCAHO) standards for patient safety issued in 2003. In order to maintain full accreditation, the JCAHO requires each health care organization to monitor for adverse events involving pharmaceuticals and devices, with medication monitoring to be a continual collaborative function. JCAHO standards indicate that medical product AE reporting should be done per applicable law/regulation, including those of state/Federal bodies.

The FDA encourages all health care providers (physicians, pharmacists, nurses, dentists, and others) to consider adverse event reporting to the FDA as part of their professional responsibility. The American Society of Health-System Pharmacists has issued guidelines on ADR monitoring and reporting.

The American Medical Association and American Dental Association advocate physician and dentist participation in adverse event reporting systems as an obligation. Since 1994, *The Journal of the American Medical Association* has instructed its authors that adverse drug or device reactions should be reported to the appropriate government agency, in addition to submitting such information for publication. The International Committee of Medical Journal Editors have revised the "Uniform Requirements for Manuscripts Submitted to Biomedical Journals" to also encourage timely reporting of urgent public health hazards.

Given the vital importance of postmarketing surveillance, MedWatch, the FDA Safety Information and Adverse Event Reporting Program, was established in 1993. While the FDA's longstanding postmarketing surveillance program predates MedWatch, this outreach initiative to health care professionals and patients was designed to promote and facilitate the voluntary reporting process by both health care providers and their patients.

The MedWatch program has four goals. The first is to increase awareness of drug, device, and other medical product-induced disease and the importance of reporting. Health professionals are taught that no drug or other medical product is without risk and are encouraged to

consider medical products as possible causes when assessing a clinical problem in a patient. This goal is accomplished through educational outreach, which includes professional presentations, publications, and a continuing education program.

The second goal of MedWatch is to clarify what should be reported. Health professionals and their patients are encouraged to limit reporting to serious AEs, enabling the FDA and the manufacturer to focus on the most potentially significant events. Causality is not a prerequisite for reporting; suspicion that a medical product may be related to a serious event is sufficient reason to notify the FDA and/or the manufacturer.

The third goal is to make it convenient and simple to submit a report of a serious AE, medication error, or product quality problem directly to the FDA. A single-page form is used for reporting suspected problems with all human-use medical products (except vaccines) regulated by the Agency—drugs, biologics, medical devices, special nutritionals (e.g., dietary supplements, medical foods, infant formulas), and cosmetics. There are two versions of the form (see Figures 7.2 and 7.3). The FDA form 3500 is used for voluntary reporting, while the FDA form 3500A is used for mandatory reporting. Both forms are available on the FDA MedWatch website (<http://www.fda.gov/medwatch>) and may be downloaded as fillable forms for saving and printing. The postage-paid FDA 3500 form may be returned to the FDA by mail or by fax to 1-800-FDA-0178.

In 1998, the MedWatch program implemented an online version of the voluntary FDA 3500 form for reporting via the Internet (see www.fda.gov/medwatch). In 2003, about 40% of the direct (voluntary) reports received from providers and consumers were sent to the FDA via this online application. In addition, MedWatch provides a toll-free 800 phone number, 1-800-FDA-1088, for reporters who wish to submit a report verbally to a MedWatch health professional.

Vaccines are the only FDA-regulated human-use medical products that are not reported on the MedWatch reporting form. Reports concerning vaccines are sent to the vaccine adverse event reporting system (VAERS) on the VAERS-1 form, available by calling 1-800-822-7967 or from the VAERS website at www.fda.gov/cber/vaers/vaers.htm. The VAERS is a joint FDA/Center for Disease Control and Prevention program for mandatory reporting by physicians of vaccine-related adverse events (see also Chapter 27).

The FDA recognizes that health professionals have concerns regarding their confidentiality as reporters, and that of the patients whose cases they report. In order to encourage reporting of adverse events, FDA regulations

offer substantial protection against disclosure of the identities of both reporters and patients. In 1995, a regulation went into effect strengthening this protection against disclosure by preempting state discovery laws regarding voluntary reports held by pharmaceutical, biological, and medical device manufacturers. In addition, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (see www.fda.gov/medwatch/hipaa.htm) specifically permits pharmacists, physicians, or hospitals to continue to report adverse events and other information related to the quality, effectiveness, and safety of FDA-regulated products (see also Chapter 19).

Manufacturers who participate in the FDA “MedWatch to Manufacturer” program (MMP) are provided with copies of serious reports submitted directly to the FDA for new molecular entities (see www.fda.gov/medwatch). To facilitate obtaining follow-up information, health professionals who report directly to the FDA are asked to indicate whether they prefer that their identity not be disclosed through the MMP to the manufacturer of the product involved in the case being reported. When such a preference is indicated, this information will not be shared.

The fourth goal of MedWatch is to provide timely and clinically useful safety information on all FDA-regulated medical products to health care professionals and their patients. The FDA’s interest in informing health professionals about new safety findings is not only to enable them to incorporate new safety information into daily practice, but also to demonstrate that voluntary reporting has a definite clinical impact.

As new information becomes available through “Dear Health Professional Letters,” public health advisories, and safety alerts, it is posted on the MedWatch website and immediate notification of the posting is sent by email to subscribers of the MedWatch listserve. This listserve reaches health care professionals, consumers, and the media. In 2004, MedWatch disseminated new safety information on over 45 drug or therapeutic biologic products as “safety alerts” to over 45 000 individual subscribers. One can subscribe to the MedWatch listserve by visiting the website (<http://www.fda.gov/medwatch/elist.htm>).

MedWatch also has a network of more than 160 health care professional, health care consumer, and health care media organizations that have allied themselves with the FDA as MedWatch Partners. Each of these organizations works with MedWatch to promote voluntary reporting and disseminate safety information notifications to their members or subscribers by using their websites, email distribution lists, and publications such as bulletins and journals.

SAFETY ASSESSMENT: THE ADVERSE EVENT REPORTING SYSTEM (AERS)

The AERS is a client-server, Oracle-based relational database system that contains all AE reports on the pharmaceuticals submitted to the Agency either directly or via the manufacturer. The mission of the AERS is to reduce adverse events related to FDA-regulated products by improving postmarketing surveillance and helping to prevent adverse outcomes related to medical errors.

The AERS was designed and implemented with the following concepts in mind:

- friendly screen layout and help function;
- enhanced search capabilities, quality control features, and electronic review of reports;
- improve the operational efficiency, effectiveness, and quality control of the process for handling AEs;
- improve the accessibility of AE information to all safety evaluators and medical officers within the FDA;
- implement and maintain compatibility with ICH standards;
- build the capability to receive electronic submissions of AEs using ICH standards;
- provide automated signal generation capabilities and improved tools for the analysis of potential AE signals.

Pharmaceutical manufacturers submit paper AE reports to the FDA central document room, where they are tracked and forwarded to the Office of Surveillance and Epidemiology (OSE) in the FDA's CDER. Reports submitted by individuals are mailed, faxed, sent via the Internet, or phoned into MedWatch, and are triaged to the appropriate FDA Center(s) (i.e., CDER, CBER, Center for Devices and Radiological Health (CDRH), Center for Veterinary Medicine (CVM), and CFSAN).

When received by the OSE, these incoming 3500 and 3500A reports are assigned a permanent report number (individual safety report), imaged, and stored in a Retrieval-Ware Imaging System; subsequently they are entered verbatim into the AERS database. Data entry has a number of sequential steps involving comparative entry, quality comparison of critical entry fields, and coding and quality control into standardized international medical terminology using MedDRA. Direct and 15-day expedited reports receive priority handling and are entered into the AERS within 14 days.

Automated quality control is performed to review reports for timeliness, completeness, and accuracy of coding. Statistical samples are also used to spot check manufacturer

performance in providing accurate and timely reports, which can be used for compliance functions.

Although the bulk of the data entry into the AERS is currently done through manual coding, the AERS is designed for electronic submission of ICH E2B(M)-standardized, MedDRA-precoded individual case safety reports. This design concept incorporates the ICH standards for content, structure, and transmittal of individual case safety reports. To prepare for full-scale implementation of electronic submissions, a step-by-step pilot program was in place. The pilot moved into full production in 2002 for capturing ICSRs.

Copies of all reports in the AERS database are available to the public through the FDA Freedom of Information Office, with all confidential information redacted (e.g., patient, reporter, institutional identifiers). The AERS database, in non-cumulative quarterly updates, can be obtained from the National Technical Information Service (www.NTIS.gov) or from the FDA website (www.fda.gov/cder/aers/extract.htm).

A variety of technology-assisted features in the AERS augment the AE review by the OSE's safety evaluators. Safety evaluators have the following pharmacovigilance tools available for AE report screening to generate signals:

- *Primary triage*: the program screens incoming reports and alerts safety evaluators to serious and unlabeled events, and serious medical events known to be drug-related (e.g., *torsade de pointes*, agranulocytosis, toxic epidermal necrolysis, etc.).
- *Secondary triage/surveillance*: provides a tool for signal identification based on overall specific counts for each risk category associated with all ADR reports received for a given drug.
- *Periodic (canned) reports*: enables periodic reviews of the AERS database, including all new actions in a time period.
- *Active (canned and/or ad hoc) query*: represents active investigation of case series signals found from any of the above levels of screening.

The AERS maximizes the ability of the Agency to identify and assess signals of importance in the spontaneous reporting system. Starting in 2006 and over a 2-year period, these upgrades will occur in what we are calling AERS II. The AERS will be upgraded to handle the FDA's processing of postmarketing adverse event reports related to human drugs and therapeutic biologics over the next 5 years. It will be web based, accept electronic submissions, meet ICH, HL7, E2B(M), eXtensible Markup Language (XML), and Tagged Image File

Format (TIFF) requirements, handle multiple product coding schemes (bar codes), interface to industry and other government systems, and include a reporting repository providing pre-tailored reports and an ad hoc feature for specialized needs.

FDA EVALUATION OF REPORTS OF ADVERSE EVENTS

Every single workday, the FDA receives nearly 1000 spontaneous reports of adverse events either directly or through the industry. The OSE in the CDER employs about 25 postmarketing safety evaluators and over a dozen epidemiologists.

The primary duty of safety evaluators is to review adverse event reports. Most of the safety evaluators are clinical pharmacists who are assigned specific groups or classes of drugs or therapeutic biologic products based on their past training and/or experience. These safety evaluators work under the tutelage and guidance of about half a dozen team leaders who have considerable experience in the evaluation and assessment of adverse event reports, substantial knowledge of the drug or therapeutic biologic agent, and awareness of the limitations of the AERS data. Every serious labeled or unlabeled adverse event report or reports describing important medical events such as liver failure, cardiac arrhythmias, renal failure, and rhabdomyolysis are electronically transferred into the computer inbox of the safety evaluators, who monitor these events daily. The safety evaluators try to identify a potential “signal,” which is defined as a previously unrecognized or unidentified serious adverse event.

Epidemiologists within the OSE are medical/clinical epidemiologists with MDs/MPHs or PhDs. Medical epidemiologists help in the “signal” development by evaluation of potential adverse event case reports (numerator data) and identification of risk factors/confounders. Epidemiologists are frequently asked to quantify and describe the exposed population (denominator data). Epidemiologists also critique published and unpublished epidemiologic studies, and participate in the design and development of protocols for epidemiologic studies submitted by drug companies in areas of regulatory interest.

The essential elements of a case report include drug name, concise description of the adverse event, date of onset of the event, drug start/stop dates if applicable, baseline patient status (comorbid conditions, use of concomitant medications, presence of risk factors), dose and frequency of administration, relevant laboratory values at baseline and during therapy, biopsy/autopsy reports, patient demographics, de-challenge (event abates when the drug is discontinued) and re-challenge

(event recurs when drug is restarted), and information about confounding drugs or conditions where available. For example, in a report describing hepatotoxicity, baseline information about liver status and information about liver enzyme monitoring would be considered essential.

If a “signal” is noted, the safety evaluator may try to find additional cases by querying the AERS database (see Case Example 7.1), doing literature searches, contacting foreign regulatory agencies directly, or collecting cases through the World Health Organization (WHO) Uppsala Monitoring Centre in Sweden (see Chapter 8). If the report is poorly documented, the safety evaluator may contact the reporter or the manufacturer for follow-up information. A case definition may be developed in collaboration with an epidemiologist and refined as new cases are identified. After a case series is assembled, the safety evaluator may look for common trends, potential risk factors, or any other items of importance. Meanwhile, with the help of drug utilization specialists in the OSE, drug usage data are obtained for the relevant drug or class of drugs or drugs within the same therapeutic category. Drug usage data are used in a variety of ways, including to obtain demographic information on the population exposed to pharmaceutical products, average duration and dose of dispensed prescription, and the specialty of the prescribing physicians. These data allow the FDA to examine how long non-hospitalized patients stay on prescription medication therapy and to learn drug combinations that may be prescribed to the same patients concurrently. These data are also used in association with AERS data to understand the context within which ADEs occur. Additionally, one or more epidemiologists may be consulted to find the background incidence of the adverse outcome in question and to estimate the reporting rates of the adverse outcome and compare it with the background rate at which the same event occurs in the population. Simply stated, a reporting rate 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. If the issue is of regulatory importance, it may be brought to the attention of others within the OSE by presentation at one of the two in-house OSE forums: the Safety Evaluator Forum and the Epi Forum. At these forums relevant personnel from the review divisions in the CDER Office of New Drugs may be invited since they are ultimately responsible for regulatory actions involving the marketing status of the product. The review division may request manufacturer-sponsored postmarketing studies to further evaluate the issue. Simultaneously, epidemiologists in the OSE may explore the feasibility of conducting pharmacoepidemiology studies in one or more large claims databases that link prescriptions with medical records. The FDA has

funded extramural researchers through a system of cooperative agreements/contracts for more than a decade. These investigators have access to large population-based databases and the FDA utilizes their resources to answer drug safety questions and to study the impact of regulatory decisions.

CASE EXAMPLE 7.1:
THE USE OF SPONTANEOUS REPORTING TO
STRENGTHEN A CONGENITAL ANOMALY
SIGNAL

Background

Efavirenz, an anti-retroviral (ARV), is indicated for the treatment of HIV/AIDS. The product, approved for marketing in 1998, is generally used in combination with other HIV/AIDS products. The prescribing information (PI) describes malformations in fetuses from efavirenz-treated monkeys that received doses that resulted in plasma concentrations similar to those in humans, including one report of anencephaly. The label recommended that women of childbearing potential undergo pregnancy testing prior to initiation of therapy and use contraception to avoid pregnancy.

Question

Can spontaneous postmarketing reports identify a signal for a rare event such as neural tube defect (NTD) in a product that is regularly used in combination with multiple other drugs, and can safety surveillance result in a regulatory decision and labeling change that supports the safe use of this product?

Approach

- The receipt of a few reports of NTDs associated with the use of efavirenz prompted a review of the spontaneous postmarketing reports in the Adverse Event Reporting System (AERS) associated with efavirenz and other marketed ARVs.
- The AERS database was searched from the date of drug approval through to November 2004. As a comparator, all other currently marketed ARVs were also searched. The case definition for NTD was based on the 2003 Guideline on NTDs from the American College of Obstetricians and Gynecologists.

Results

- A total of 128 reports were retrieved from the AERS, including many reports from the Antiretroviral

Pregnancy Registry. After merging duplicates and excluding reports where the reported event did not meet the definition for an NTD or the time of maternal drug exposure was not reasonably associated with fetal neural tube development, nine cases were identified: five with efavirenz, four without efavirenz.

- When the proportion of NTDs among the full range of reported congenital anomalies was compared for each of the marketed ARV medications, there was a disproportionately higher percentage of efavirenz cases reporting NTDs. Among all reported congenital anomalies, NTDs were about three times more frequently reported for efavirenz than any of the other ARVs. The presence of these NTD reports in the AERS supports the safety signal identified in animal studies.

Outcome

- Voluntary adverse event reporting and review of spontaneously reported adverse events resulted in identification of a serious adverse outcome not identified in clinical trials in humans.
- The Pregnancy Category was changed from C (animal studies showed an adverse effect on the fetus, but no adequate and well-controlled studies in humans) to D (positive evidence of human fetal risk based on adverse reaction data from investigational or marketing experience or studies in humans).
- The Warnings, Precautions/Pregnancy and Information for Patients, and Patient Information sections of the prescribing information were updated to inform health care professionals and patients about this new safety information. The FDA disseminated this labeling information and the manufacturer's Dear Healthcare Professional letter on the FDA MedWatch website at: <http://www.fda.gov/medwatch/SAFETY/2005/safety05.htm#Sustiva>.

Strengths

- A separate prospective ARV pregnancy registry provided adequate case information to assess the timing of fetal drug exposure and allow evaluation of risk factors for NTDs.
- Disproportionate reporting of this type of congenital anomaly identified that efavirenz was unique among 19 ARVs, some with over 15 years of postmarketing experience.

Limitations

Because of a lack of numerator (number of cases) and denominator (total number of patients exposed to the drug) for the AERS data, it is not possible to determine the incidence of a particular event from postmarketing data.

Summary Points

- Spontaneous reporting of drug adverse events may allow for identification and evaluation of rare drug-associated events, even when the products are used in combination with multiple other products.
- Spontaneous reporting of drug adverse events may allow for comparisons across drug classes, even when products have very different numbers of reports in the AERS database. Because of incomplete reporting and reporting bias, it is important to approach such comparisons with caution.

After confirmation of a “signal” the FDA can initiate various regulatory actions, the extent and rigor of which depend on the seriousness of the adverse event, the availability, safety, the acceptability of alternative therapy, and the outcome of previous regulatory interventions. Regulatory interventions to manage the risk include labeling change such as a boxed warning, restricted use or distribution of the drug, name, or packaging change(s), a “Dear Health Care Professional” letter, or, rarely, possible withdrawal of a medical product from the market (see Table 7.3 and also Chapter 27).

Table 7.3. Recent safety-based drug withdrawals

Drug name	Year approved/year withdrawn
Phenylpropanolamine	—/2000 (never approved by FDA)
Fenfluramine	1973/1997
Terfenadine	1985/1998
Astemizole	1988/1999
Cisapride	1993/2000
Dexfenfluramine	1996/1997 (not an NME) ^a
Bromfenac	1997/1998
Cerivastatin	1997/2001
Grepafloxin	1997/1999
Mibefradil	1997/1998
Troglitazone	1997/2000
Rapacuronium	1999/2001
Rofecoxib	1999/2004
Alosetron ^b	2000/2000
Valdecoxib	2001/2005

^a NME, new molecular entity.

^b Returned to market in 2002 with restricted distribution.

The time between the first identification of a safety risk and the implementation of a regulatory action may take several months to years depending on the nature of the problem and the public health impact. For example, several years elapsed between the time when dangerous drug interactions with cisapride and a number of other drugs were identified and when the drug was ultimately removed from the market for general use. Similarly, severe liver failure in association with the use of the antidiabetic drug troglitazone was noted a few months after marketing but it took a few years before the drug was removed from the market. In the examples of both cisapride and troglitazone, a variety of regulatory interventions, such as repeated labeling changes and “Dear Health Care Professional” letters, were applied over the years to manage the risk before these products were removed from the market. These regulatory interventions did not achieve meaningful improvement in prevention of contraindicated drug use or in liver enzyme testing, respectively.

To notify health professionals of important new safety information discovered after marketing, the FDA often requests that the manufacturer send a “Dear Health Care Professional” letter to warn providers of particular safety issues. This is done in combination with a labeling change, although only a small proportion of labeling changes result in such letters. Frequently, the change in labeling may be accompanied by issuance of a press release (also known as a Talk Paper) or public health advisory. Additionally, FDA scientists may disseminate new drug safety information through publications in professional journals and presentations at professional meetings.

There were 43 drug or biologic letters/safety notifications posted in 2002 and 36 in 2003. In 2003, safety-related labeling changes were approved by the FDA for 20–45 drug products each month. “Dear Health Care Professional” letters and other safety notifications, and summaries of safety-related labeling changes approved each month, can be found on the MedWatch website (www.fda.gov/medwatch/safety.htm). Table 7.4 lists some examples of recent “Dear Health Care Professional” letters.

The FDA can seek to restrict or limit the use of a drug product through labeling if the adverse reaction associated with the drug has severe consequences. Foreexample, the labeling for the new arthritis/pain drug valdecoxib was strengthened with new warnings following postmarketing reports of serious adverse effects, including life-threatening risks related to skin reactions (e.g., Stevens–Johnson Syndrome and anaphylactoid reactions). The labeling advised people who started valdecoxib and experienced

Table 7.4. Recent FDA MedWatch safety alerts/"Dear Health Care Professional" letters, 2003

Drug	Details
Topamax® (topiramate)	Revised the WARNINGS and PRECAUTIONS to notify health care professionals that Topamax causes hyperchloremic, non-anion gap metabolic acidosis (decreased serum bicarbonate). Measurement of baseline and periodic serum bicarbonate during topiramate treatment is recommended.
Permax® (pergolide mesylate)	Revised the WARNINGS and PRECAUTIONS sections to inform health care professionals of the possibility of patients falling asleep while performing daily activities, including operation of motor vehicles, while receiving treatment with Permax®. Many patients who have fallen asleep have perceived no warning of somnolence.
Arava® (leflunomide)	In postmarketing experience worldwide, rare, serious hepatic injury, including cases with fatal outcome, have been reported during treatment with Arava. Most cases occurred within six months of therapy and in a setting of multiple risk factors for hepatotoxicity.
Viread® (tenofovir disoproxil fumarate)	Notified health care professionals of a high rate of early virologic failure and emergence of nucleoside reverse transcriptase inhibitor resistance associated mutations in a clinical study of HIV-infected treatment-naïve patients receiving a triple regimen of didanosine, lamivudine and tenofovir disoproxil fumarate.
Lariam® (mefloquine hydrochloride)	Notified health care professionals of the Lariam Medication Guide developed in collaboration with the FDA to help travelers better understand the risks of malaria, the risks and benefits associated with taking Lariam to prevent malaria, and the potentially serious psychiatric adverse events associated with use of the drug.
Prandin® (repaglinide)	Revised the PRECAUTIONS/Drug Interaction section to inform health care professionals of a drug–drug interaction between repaglinide and gemfibrozil. Concomitant use may result in enhanced and prolonged blood glucose-lowering effects of repaglinide.
Serevent Inhalation Aerosol® (salmeterol xinafoate)	New labeling includes a boxed warning about a small, but significant, increased risk of life-threatening asthma episodes or asthma-related deaths observed in patients taking salmeterol in a recently completed large US safety study.
Ziagen® (abacavir)	High rate of early virologic non-response observed in a clinical study of therapy-naïve adults with HIV infection receiving once-daily three-drug combination therapy with lamivudine (Epivir, GSK), abacavir (Ziagen, GSK), and tenofovir (Viread, TDF, Gilead Sciences).
Genotropin® (somatropin [rDNA origin] for injection)	Fatalities have been reported with the use of growth hormone in pediatric patients with Prader–Willi syndrome with one or more of the following risk factors: severe obesity, history of respiratory impairment or sleep apnea, or unidentified respiratory infection.
Topamax® (topiramate) tablets/sprinkle capsules	Oligohidrosis (decreased sweating) and hyperthermia have been reported in topiramate-treated patients. Oligohidrosis and hyperthermia may have potentially serious sequelae, which may be preventable by prompt recognition of symptoms and appropriate treatment.
Risperdal® (risperidone)	Cerebrovascular adverse events (e.g., stroke, transient ischemic attack), including fatalities, were reported in patients in trials of risperidone in elderly patients with dementia-related psychosis.
Avonex® (Interferon β-1a)	Postmarketing reports of depression, suicidal ideation and/or development of new or worsening of pre-existing psychiatric disorders, including psychosis, and reports of anaphylaxis, pancytopenia, thrombocytopenia, autoimmune disorders of multiple target organs, and hepatic injury.

a rash to discontinue the drug immediately and also the drug was contraindicated in patients allergic to sulfa-containing products. The drug has recently been removed from the market.

Drug safety problems can also lead to the removal of a drug from the market (see Case Example 7.2). Fortunately, such product withdrawals are very uncommon; there have been only 22 drugs taken off the US market since 1980; drugs withdrawn recently are listed in Table 7.3.

CASE EXAMPLE 7.2:
SIGNAL IDENTIFIED VIA SPONTANEOUS
REPORTING CONFIRMED BY A FORMAL
PHARMACOEPIDEMIOLOGY STUDY

Background

- Phenylpropanolamine (PPA) is an ingredient that was used in many over-the-counter (OTC) and prescription

cough and cold medications as a decongestant and in OTC weight loss products. As early as 1984, the FDA received reports of hemorrhagic stroke (bleeding into the brain or into tissue surrounding the brain) in association with PPA. In addition there were published reports in the literature.

Question

- Is the use of PPA-containing products associated with hemorrhagic stroke?

Approach

- To confirm the signal of hemorrhagic stroke in association with PPA use identified through case reports submitted to the FDA's spontaneous reporting system database, and published case reports in the literature, an ad hoc case-control study was conducted by researchers at Yale University (Kernan WN, *et al.* Phenylpropanolamine and the risk of hemorrhagic stroke. *N Engl J Med* 2000; **343**: 1826–32).

Results

- The study demonstrated a statistically significant increased risk of hemorrhagic stroke among both appetite suppressant users and first-time users of PPA as a cough/cold remedy.

Outcome

- An FDA Advisory Committee meeting was convened to discuss the case-control study, along with additional information on PPA, and determined that there is an association between PPA and hemorrhagic stroke and recommended that PPA be considered not safe for OTC use.
- The FDA agreed with the recommendations of its Advisory Panel and took steps to remove PPA from all drug products and requested all drug companies to discontinue or reformulate PPA-containing products.

Strength

- The case-control study is a good example of collaboration between scientists at the FDA and academic researchers.

Limitation

- The case-control study was a major undertaking and required more than 5 years to complete.

Summary Points

- A formal epidemiological study is usually needed to confirm a signal identified through spontaneous reports.

In addition to the technology used in current adverse event reporting, including sophisticated relational databases and network connections for electronic transfer, new methods to evaluate and assess spontaneous reports are being explored to take advantage of the sheer volume of data. Aggregate analysis tools and data mining techniques are currently being developed by the OSE, WHO, and others, to systematically screen large databases of spontaneous reports.

Since 1998, the FDA has explored automated and rapid Bayesian data mining techniques to enhance its ability to monitor the safety of drugs, biologics, and vaccines after they have been approved for use. In May 2003, the FDA announced the establishment of a Cooperative Research and Development Agreement (CRADA) with a private software development company. The CRADA is expected to improve the utility of safety data mining technology. The FDA's CDER and CBER will work with this private company to develop new and innovative ways for extracting information related to drug safety and risk assessment. To this end, a desktop data mining software tool, called WebVDME, has been developed and is currently being piloted.

Data mining is a technique for extracting meaningful, organized information from large complex databases. In data mining the strategy is to use a computer to identify potential signals in large databases that might be overlooked, for a variety of reasons, in a manual review on a case-by-case basis. Drug-AE signals are generated by comparing the frequency of reports with what would be expected if all drugs and AEs were assumed to follow certain patterns. The goal is to distinguish the more important or stronger signals to facilitate identification of combinations of drugs and events that warrant more in-depth follow-up. Data mining is a tool best suited for generation of possible signals and it cannot replace or override the meticulous hands-on review by safety evaluators. Further, whether it has any advantage over the hands-on review, and the degree to which it generates false signals, remains to be evaluated.

STRENGTHS

LARGE-SCALE AND COST-EFFECTIVE

Two vital advantages of surveillance systems based on spontaneous reports are that they potentially maintain ongoing surveillance of all patients, and are relatively inexpensive. Spontaneous reporting systems are the most common method used in pharmacovigilance to generate signals on new or rare adverse events not discovered during clinical trials.

GENERATION OF HYPOTHESES AND SIGNALS

Making the best possible use of the data obtained through monitoring underlies postmarketing surveillance. Toward that goal, the great utility of spontaneous reports lies in *hypothesis generation*, with need to explore possible explanations for the adverse event in question. By raising suspicions, spontaneous report-based surveillance programs perform an important function, which is to generate *signals* of potential problems that warrant further investigation.

Assessment of the medical product–adverse event relationship for a particular report or series of reports can be quite difficult. Table 7.5 lists factors that are helpful in evaluating the strength of association between a drug and a reported adverse event. (See also Chapter 17.)

The stronger the drug–event relationship in each case and the lower the incidence of the adverse event occurring spontaneously, the fewer case reports are needed to perceive causality. It has been found that for rare events, coincidental drug–event associations are so unlikely that they merit little concern, with greater than three reports constituting a signal requiring further study. In fact, it has been suggested that a temporal relationship between medical product and adverse event, coupled with positive de-challenge and re-challenge, can occasionally make

isolated reports conclusive as to a product–event association. Biological plausibility and reasonable strength of association aid in deeming any association as causal (see also Chapter 17).

However, achieving certain proof of causality through adverse event reporting is unusual. Confirmation of an association between a drug and an adverse reaction usually requires further additional studies. Attaining a prominent degree of suspicion is much more likely, but still may be considered a sufficient basis for regulatory decisions.

OPPORTUNITY FOR CLINICIAN CONTRIBUTIONS

The reliance of postmarketing surveillance systems on health professional reporting enables an individual to help improve public health. This is demonstrated by one study that found direct practitioner participation in the FDA spontaneous reporting system to be the most effective source of new ADR reports that led to changes in labeling. Ensuring that the information provided in the adverse event report is as complete and in-depth as possible further enhances postmarketing surveillance. Even consumers can have an impact on detecting signals and leading to regulatory changes (see Case Example 7.3). Thus, while possessing inherent limitations, postmarketing surveillance based on spontaneous reports data is a powerful tool for detecting adverse event signals of direct clinical impact.

CASE EXAMPLE 7.3: **THE VALUE OF CONSUMER REPORTING** **OF ADVERSE DRUG EVENTS**

Background

- Mitoxantrone, an antineoplastic agent, is also indicated for reducing neurologic disability and/or the frequency of clinical relapses in patients with certain forms of multiple sclerosis (MS), which is believed to be an autoimmune disease. The product label of mitoxantrone contains a black box warning (FDA's strongest form of warning for medicines) for cardiac (heart) toxicity, which can occur either during therapy or months to years after termination of therapy with mitoxantrone.

Table 7.5. Useful factors for assessing causal relationship between drug and reported adverse events

-
- Chronology of administration of agent, including beginning and ending of treatment and adverse event onset
 - Course of adverse event when suspected agent stopped (de-challenge) or continued
 - Etiologic roles of agents and diseases in regard to adverse event
 - Response to readministration (re-challenge) of agent
 - Laboratory test results
 - Previously known toxicity of agent
-

Question

- Can a consumer report of a serious labeled adverse event (cardiac toxicity in this case) result in additional regulatory action by the FDA?

Approach

- Receipt of a consumer report of severe heart failure in a young mitoxantrone-treated MS patient who had to undergo a heart transplant prompted a review of the FDA's Adverse Event Reporting System (AERS) for additional cases of cardiac toxicity.

Results

- A search in the AERS yielded 38 unduplicated reports of serious cardiac toxicity in association with mitoxantrone use.
- Cases of cardiac toxicity were also noted in patients who were administered cumulative doses below the current monitoring threshold.

Outcome

- Evidence from spontaneous reports submitted to the FDA, together with published case reports and data from an interim analysis of a postmarketing study in MS patients, resulted in a revised warning for cardiac toxicity in association with mitoxantrone use.
- An educational program was initiated, directed to physicians (Dear Healthcare Professional letter) and patients (revised Patient Information sheet), together with a MedWatch safety alert regarding cardiac toxicity in association with mitoxantrone use.

Strength

- A spontaneous reporting system is a relatively inexpensive method to identify cases of drug-associated adverse events.

Limitation

- Because of lack of numerator (number of cases) and denominator (total number of patients exposed to the drug) data, it is not possible to determine the incidence of a particular event from the AERS database.

Summary Points

- Reporting of adverse drug experiences that are already labeled is important and may result in regulatory action.
- Reporting of serious adverse events by consumers can help postmarketing surveillance.

WEAKNESSES

There are important limitations to consider when using spontaneously reported adverse event information. These limitations include difficulties with adverse event recognition, underreporting, biases, estimation of population exposure, and report quality.

ADVERSE EVENT RECOGNITION

The attribution of AEs (or any other medical product-associated adverse event) may be quite subjective and imprecise. While an attribution of association between the medical product and the observed event is assumed by the reporters with all spontaneously reported events, every effort is made to rule out other explanations for the event in question. It is well known that placebos and even no treatment can be associated with adverse events. In addition, there is almost always an underlying background rate for any clinical event in a population, regardless of whether there was exposure to a medical product.

Reaching a firm conclusion about the relationship between exposure to a medical product and the occurrence of an adverse event can be difficult. In one study, clinical pharmacologists and treating physicians showed complete agreement less than half the time when determining whether medication, alcohol, or "recreational" drug use had caused hospitalization.

Such considerations emphasize the crucial need for careful, thoughtful review of adverse event reports upon their receipt by the FDA or the manufacturer. It is through this process that causality, or at least a high degree of suspicion for a product-adverse event association, is put to the test (see also Chapter 17). Ultimately, formal pharmacoepidemiology studies are usually needed to strengthen the observed association.

UNDERREPORTING

Another major concern with any spontaneous reporting system is underreporting of adverse events. The extent of underreporting is unknown and may be influenced by the severity of the event, the specialty of the reporter, how

long the drug has been on the market, whether the event is labeled, and whether the drug is prescription or non-prescription. It has been estimated that rarely more than 10% of serious ADRs, and 2–4% of non-serious reactions, are reported to the British spontaneous reporting program. A similar estimate is that the FDA receives by direct report less than 1% of suspected serious ADRs. This means that cases spontaneously reported to any surveillance program, which comprise the numerator, generally represent only a small portion of the number that have actually occurred. The impact of underreporting can be somewhat lessened if submitted reports, irrespective of number, are of high quality.

BIASES

Spontaneously reported information is subject to the influence of a number of biases. These include the length of time a product has been on the market, size of sponsors' detail force, target population, health care providers' awareness, the quality of the data, and publicity effects.

In addition, it has been observed that spontaneous reporting of adverse events for a drug tends to peak at the end of the second year of marketing and then declines thereafter (Weber effect).

In addition to these biases, it is possible that reported cases might differ from nonreported cases in characteristics such as time to onset or severity.

ESTIMATION OF POPULATION EXPOSURE

Compounding these limitations is the lack of denominator data, such as user population and drug exposure patterns, that would provide an estimate of the number of patients exposed to the medical product, and thus at risk for the adverse event of interest. Numerator and denominator limitations make incidence rates computed from spontaneously reported data problematic, if not completely baseless. However, even if the exposed patient population is not precisely known, estimation of the exposure can be attempted through the use of drug utilization data.

This approach, whose basic methodologies are applicable to medical products in general, can be of utility. Major sources of data on the use of drugs by a defined population include market surveys based on sales or prescription data, third-party payers or health maintenance organizations, institutional/ambulatory settings, or specific pharmacoepidemiology studies (see Section III). Cooperative agreements and contracts with outside researchers enable the FDA to use such databases in its investigations. Care must be taken in

interpreting results from studies using these databases. That drug prescribing does not necessarily equal drug usage, and the applicability of results derived from a specific population (such as Medicaid recipients) to the population at large, need to be weighed carefully.

REPORT QUALITY

The ability to assess, analyze, and act on safety issues based on spontaneous reporting is dependent on the quality of information submitted by health professionals in their reports. A complete adverse event report should include the following:

- product name (and information such as model and serial numbers in the case of medical devices);
- demographic data;
- succinct clinical description of the adverse event, including confirmatory/relevant test/laboratory results;
- confounding factors such as concomitant medical products and medical history;
- temporal information, including the date of event onset and start/stop dates for use of medical product;
- dose/frequency of use;
- biopsy/autopsy results;
- de-challenge/re-challenge information;
- outcome.

SUMMARY

The major limitations of the FDA's AE reporting system reflect the fact that the data are generated in an uncontrolled and incomplete manner. Although manufacturers are legally required to submit AE reports to the FDA and some of those reports are based on formal studies, the majority of AEs originate with practicing physicians who may or may not notify the manufacturer or the FDA when they observe an AE in one of their patients. It appears that they generally do not choose to report AEs, and the number of reports that the FDA receives is not representative of the extent of adverse events that occur in the United States. The number of reports in the system is also influenced by a variety of other factors, such as the extent and quality of the individual manufacturer's postmarketing surveillance activities, the nature of the event, the type of drug, the length of time it has been marketed, and publicity in the lay or professional press. Because of these limitations and the others outlined

in this chapter, AE reports are primarily useful for hypothesis generating, rather than hypothesis testing. Ironically, the scientifically uncontrolled nature of AE reporting creates its greatest advantage—the ability to detect and characterize AEs occurring across a broad range of medical practice—as well as its most serious limitations.

PARTICULAR APPLICATIONS

OVERALL

The FDA's AERS contains almost 3 million reports, with the earliest dating back to 1969. While reporting levels remained fairly constant during the 1970s—about 18 000 reports were entered into the database in 1970, and slightly over 14 000 reports were added in 1980—reporting increased dramatically after 1992, as can be seen in Figure 7.4. By 1992, the annual number of reports had risen to 120 000, and in 2003 was over 370 000. Forty percent of these reports were serious and unexpected (i.e., 15-day).

As noted earlier, the AERS contains reports from a variety of sources. Reports may be from the United States or other countries. The suspected AEs may have been observed in the usual practice of medicine or during formal studies; case reports from the literature are also included. Reports come to the FDA either directly from health professionals or consumers, or from pharmaceutical manufacturers. The vast majority (over 90%) of adverse drug event reports are received by the FDA through the manufacturer, with the remainder received directly from health care professionals or consumers.

In 2003, of all voluntary reports sent directly to the FDA, 68% involved drugs, 14% medical devices, 12% drug quality problems, 3% biologics, and 3% dietary supplements. The sources were: 59% from pharmacists, 15% from physicians, 9% from nurses, and 6% from non-health professionals (with 11% source not given).

SPECIFIC EXAMPLES

Temafloxacin (Omniflox®): Withdrawn from Market

This oral antibiotic was first marketed in February 1992. During the first three months of its use, the FDA received approximately 50 reports of serious adverse events, including three deaths. These events included hypoglycemia in elderly patients as well as a constellation of multi-system organ involvement characterized by hemolytic anemia,

frequently associated with renal failure, markedly abnormal liver function tests, and coagulopathy. When approved by the FDA, temafloxacin was already being used in Argentina, Germany, Italy, Ireland, Sweden, and the United Kingdom. However, the FDA's experience with this drug demonstrates the critical importance of postmarketing surveillance and the timely reporting of adverse events. Prior to FDA approval, slightly more than 4000 patients had received the drug in clinical trials, and temafloxacin was considered to have a side effect profile similar to other quinolone antibiotics. In its first three months of commercial marketing, many thousands of patients received the drug. Only after this much broader clinical experience did the serious side effects described above become apparent. Less than four months after its introduction into the marketplace, the drug was withdrawn.

Linezolid (Zyvox®): Serious, Unlabeled ADR Noted Shortly After Approval

Linezolid (Zyvox®), a synthetic antibacterial agent of the oxazolidinone class, was approved for use in April 2000. It is indicated for the treatment of adult patients with the following infections caused by susceptible strains of designated microorganisms: vancomycin-resistant *Enterococcus faecium*, including cases with concurrent bacteremia; nosocomial pneumonia; complicated and uncomplicated skin and skin structure infections; and community-acquired pneumonia, including cases with concurrent bacteremia.

At the time of approval, safety data were limited, based primarily on its use in controlled clinical trials. The most serious adverse event noted in the initial product labeling was thrombocytopenia, mentioned in the Precautions section and the Laboratory Changes subsection of the Adverse Reactions section.

As reported in the Animal Pharmacology section of the product labeling, linezolid had caused dose- and time-dependent myelosuppression, as evidenced by bone marrow hypocellularity, decreased hematopoiesis, and decreased levels of circulating erythrocytes, leukocytes, and platelets in animal studies.

Within the first six months the drug was on the market, four cases of red cell aplasia associated with its use were received by the FDA. In addition, six other cases suggestive of myelosuppression had been submitted, as well as two cases of sideroblastic anemia.

With the increasing number of cases being received by the FDA, an in-depth review of this problem was undertaken. The AERS was searched for reports of hematologic

toxicity associated with linezolid and a total of 27 reports were retrieved through to September 20, 2000. These reports were reviewed to find any that may have been suggestive of myelosuppression but were not necessarily reported as such (e.g., reductions in white blood count, hemoglobin and hematocrit, and platelets). In addition to the four red cell aplasia cases, six additional cases suggestive of myelosuppression were identified:

- A bone marrow transplant recipient who had a delayed engraftment that was thought to be due to linezolid myelosuppression.
- Three cases reported as routine complete blood counts (CBC), revealing decreased white blood cells (WBC), hemoglobin and hematocrit, and platelets. Personal communication with the reporters in these three cases found no further follow-up such as bone marrow biopsy, nor progression to more serious disease.
- Two cases were received as direct reports: one described as bone marrow suppression and thrombocytopenia in a 65-year-old male and the other as pancytopenia in a 51-year-old female.

Because of the rapidity with which these cases were reported to the FDA in the short time linezolid had been on the market, and the relatively small estimated number of courses of therapy sold, the FDA and the manufacturer agreed to prominent warnings to be included in the labeling concerning the development of myelosuppression. Changes were made to the Warnings and Precautions sections to recommend to clinicians that:

Myelosuppression (including anemia, leukopenia, pancytopenia, and thrombocytopenia) has been reported in patients receiving linezolid. In cases where the outcome is known, when linezolid was discontinued, the affected hematologic parameters have risen toward pretreatment levels. Complete blood counts should be monitored weekly in patients who receive linezolid, particularly in those who receive linezolid for longer than two weeks, those with pre-existing myelosuppression, those receiving concomitant drugs that produce bone marrow suppression, or those with a chronic infection who have received previous or concomitant antibiotic therapy. Discontinuation of therapy with linezolid should be considered in patients who develop or have worsening myelosuppression.

Valproic Acid (Depakote®): Increased Severity of Labeled ADR Noted After Many Years of Use

Valproic acid products, including Depakote®, Depakene®, and Depacon®, have been used in clinical care since FDA

approval in 1978. Although pancreatitis was first listed in the package inserts of valproate products in 1981, as with most drugs there was limited safety data on this product at the time of approval. In clinical trials, there were two cases of pancreatitis without alternative etiology in 2416 patients, representing 1044 patient-years experience.

Initially, these drugs were indicated for a narrow labeled use and a limited population. Over two decades, the product was used for a wider range of both on-label and off-label indications, and the population exposed to the drug included a broader population than that exposed during the pre-approval clinical trials. With this increased use, the FDA received a number of voluntary reports through the MedWatch spontaneous reporting system of more severe forms of pancreatitis, often hemorrhagic, sometimes fatal, and with a number of cases occurring in infants and adolescent children. Although this ADR, pancreatitis, was “labeled” or known, the increased severity of the condition prompted the OSE postmarketing surveillance staff and the review division to initiate an epidemiological investigation and the development of a case series. This evaluation demonstrated that the rate based upon the reported cases exceeded that expected in the general population and there were cases in which pancreatitis recurred after re-challenge with valproate. With the agreement of the manufacturer, the FDA approved new safety labeling changes to the Warnings and Precautions sections and modified a black box warning to inform clinicians and their patients:

Pancreatitis: cases of life-threatening pancreatitis have been reported in both children and adults receiving valproate. Some of the cases have been described as hemorrhagic with rapid progression from initial symptoms to death. Cases have been reported shortly after initial use as well as after several years of use. Patients and guardians should be warned that abdominal pain, nausea, vomiting, and/or anorexia can be symptoms of pancreatitis that require prompt medical evaluation. If pancreatitis is diagnosed, valproate should ordinarily be discontinued. Alternate treatment for the underlying medical condition should be initiated as clinically indicated (see warnings and precautions).

THE FUTURE

The systematic collection and evaluation of postmarketing reports of serious ADRs by the FDA has come a long way since its inception about 50 years ago. The May 1999 report to the FDA Commissioner, *Managing the Risks from Medical Product Use: Creating a Risk Management Framework*, found that the postmarketing surveillance program currently

in place performed well for the goal it was designed to achieve—the rapid detection of unexpected serious AEs. Yet, it should be remembered that spontaneous reporting, although invaluable, is only one tool used in managing medical product risk. The report recognized that the FDA's programs are not designed to evaluate the rate, or impact, of known adverse events. The report proposed several options for improving risk management, including expanding the use of automated systems for reporting, monitoring, and evaluating AEs, and increasing the Agency's access to data sources that would supplement and extend its spontaneous reporting system. This could include use of large-scale medical databases from health maintenance organizations to reinforce, support, and enhance spontaneous signals and provide background rates and descriptive epidemiology.

Since the 1999 report, the FDA has continued to work with academia and industry to address these recommendations. In recognition of the increasing importance of postmarketing surveillance and risk assessment in the regulatory setting, a variety of initiatives are under way within the FDA. In 2002, the OSE was created within the CDER, with its three divisions focusing on improved identification and epidemiologic evaluation of ADRs, the evaluation of medication errors, and further research and implementation of risk communication activities directed toward both health care professionals and patients. The recent reauthorization of the Prescription Drug Users Fee Act (PDUFA) in 2002 will, for the first time, allow the FDA to apply user fee funds to the postmarketing activities of the Agency. In anticipation of these expanded efforts the FDA has published several guidance documents on postmarketing risk evaluation, risk communication, and risk management (see www.fda.gov/bbs/topics/news/2004/NEW01059.html).

In 2003, the OSE initiated a formal, competitive process of direct access to longitudinal, patient-level, electronic medical record data which can be used to study ADRs. Acquisition of this resource will directly enhance the OSE's ability to achieve one of the FDA's strategic goals, i.e., improving patient and consumer safety. In addition, online access to this data resource will allow the OSE to conduct drug safety studies in large population-based settings.

The FDA's current and future efforts include the following: increasing the quality of incoming reports of adverse events with a focus on making the AERS more efficient; establishing global reporting standards; promoting speed of reporting and assessment through electronic reporting; exploring new assessment and data visualizing methodologies; and, finally, exploring tools beyond spontaneous reporting. The last initiatives involve identification and

assessment of linked databases and registries which can be accessed to expand surveillance, provide confirmatory evidence for signals, assess regulatory impact of labeling changes through studies, and, in general, build on the known strengths of spontaneous reporting—signal generation of potentially important events.

In addition, the OSE will refine current techniques to assess drug risks through the development and evaluation of risk management programs. We will continue to consider appropriate risk communication tools in order to clearly articulate drug safety information to both health professionals and patients in a timely manner. Our goals for the next 3–5 years include plans to develop and establish “best practices” for risk management plans and to develop quantitative approaches to the review of postmarketing safety data.

In summary, spontaneous reporting of AEs provides an important cornerstone for pharmacovigilance in the US. Regulators and manufacturers of medical products worldwide are moving forward the “single safety message transmission” with global harmonization for data standards and data transmission, improvements in relational database systems, the development of new risk assessment methodologies, and increased access to other data resources, including computerized medical records, to improve our overall ability to manage risk from pharmaceuticals.

DISCLAIMER

The opinions expressed are those of the authors and do not necessarily represent the views of the FDA or the US Government.

Key Points

- Spontaneous reports of adverse events (AEs) are sent to the FDA either directly through the MedWatch program or indirectly via the manufacturer.
- In the US, reporting of AEs is required by law for manufacturers but is voluntary for health care providers and consumers.
- In addition to hands-on review of AE reports by safety evaluators, the FDA employs data mining techniques to identify signals (previously unrecognized or unidentified serious AE).
- In spite of its limitations (e.g., underreporting; incomplete reports), spontaneous reports of AEs provide an important cornerstone for pharmacovigilance in the US.

SUGGESTED FURTHER READINGS

- Ahmad SR, Freiman JP, Graham DJ, Nelson RC. Quality of adverse drug experience reports submitted by pharmacists and physicians to the FDA. *Pharmacoepidemiol Drug Saf* 1996; **5**: 1–7.
- Ahmad SR. Adverse drug event monitoring at the Food and Drug Administration. *J Gen Intern Med* 2003; **18**: 57–60.
- Bonnel RA, Graham DJ. Peripheral neuropathy in patients treated with leflunomide. *Clin Pharm Therapeutics* 2004; **75**: 580–5.
- Brinker A, Staffa J. Concurrent use of selected agents with moxifloxacin: an examination of labeling compliance within 1 year of marketing. *Arch Intern Med* 2002; **162**: 2011–12.
- FDA. *Managing the Risks from Medical Product Use*. Report to the FDA Commissioner from the Task Force on Risk Management. Rockville, MD: Food and Drug Administration, May 1999. Available at: <http://www.fda.gov/oc/tfrm/riskmanagement.pdf/>.
- Flowers C, Racoosin J, Lu S, Beitz J. Pergolide-associated valvular heart disease. *Mayo Clin Proc* 2003; **78**: 730–1.
- Graham DJ, Drinkard CR, Shatin D, Tsong Y, Burgess MJ. Liver enzyme monitoring in patients treated with troglitazone. *JAMA* 2001; **286**: 831–3.
- Graham DJ, Drinkard CR, Shatin D. Incidence of idiopathic acute liver failure and hospitalized liver injury in patients treated with troglitazone. *Am J Gastroenterol* 2003; **98**: 175–9.
- Kessler DA. Introducing MedWatch: a new approach to reporting medication and device adverse effects and product problems. *JAMA* 1993; **269**: 2765–8.
- Rossi AC, Hsu JP, Faich GA. Ulcerogenicity of piroxicam: analysis of spontaneously reported data. *BMJ* 1987; **294**: 147–50.
- 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**: 311–16.
- Shaffer D, Singer S, Korvick J, Honig P. Concomitant risk factors in reports of torsades de pointes associated with macrolide use: review of the United States Food and Drug Administration Adverse Event Reporting System. *Clin Infect Dis* 2002; **35**: 197–200.
- 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.
- Staffa JA, Chang J, Green L. Cerivastatin and reports of fatal rhabdomyolysis. *N Engl J Med* 2002; **346**: 539–40.
- Tsong Y. Comparing reporting rates of adverse events between drugs with adjustment for year of marketing and secular trends in total reporting. *J Biopharm Stat* 1995; **5**: 95–114.
- Wysowski DK, Farinas E, Swartz L. Comparison of reported and expected deaths in sildenafil (Viagra) users. *Am J Cardiol* 2002; **89**: 1331–4.
- Wysowski DK, Honig SF, Beitz J. Uterine sarcoma associated with tamoxifen use. *N Engl J Med* 2002; **346**: 1832–3.

8

Global Drug Surveillance: The WHO Programme for International Drug Monitoring

The following individuals contributed to editing sections of this chapter:

I. RALPH EDWARDS, STEN OLSSON, MARIE LINDQUIST, and BRUCE HUGMAN
WHO Collaborating Centre for International Drug Monitoring (Uppsala Monitoring Centre), Uppsala, Sweden.

INTRODUCTION

The general awareness that modern drugs could carry unexpected hazards was triggered by a letter to the editor of the *Lancet* published on December 16, 1961. In this historical document of fifteen lines, Dr McBride from Australia reported that he had noted an increased frequency of limb malformations among babies, and that a common denominator seemed to be the intake of a new hypnotic drug—thalidomide—by their mothers.

In the wake of the public health disaster that then unraveled, governments in many countries arranged procedures for the systematic collection of information about suspected adverse drug reactions (ADRs). These systems were based on the spontaneous reporting of suspected ADRs by physicians. They were first organized in Australia, Canada, Czechoslovakia, Ireland, the Netherlands, New Zealand, Sweden, the UK, the US, and West Germany. They were initiated between 1961 and 1965. Similar systems now operate in more than 70 countries. Many of the principles which

are still important in pharmacovigilance were elaborated in these early days, mainly by Finney.

In 1968, ten countries from Australasia, Europe, and North America agreed to pool all reports that had been sent to their national monitoring centers in a WHO-sponsored international drug monitoring project. The aim was to identify even very rare but serious reactions as early as possible. The scheme was set up at WHO headquarters in Geneva in 1970. The economic and operational responsibilities were transferred to Sweden in 1978 with the establishment of the WHO Collaborating Centre for International Drug Monitoring in Uppsala (now known as the Uppsala Monitoring Center, UMC). The formal responsibility for and the coordination of the program, however, remained with WHO headquarters. Today, 73 countries participate in the program as full members and a further 12 as associate members (Figure 8.1), annually contributing approximately 200 000 suspected ADR reports to the WHO database in Uppsala. This database holds nearly three million case reports to date. There are guidelines covering all aspects of reporting, and



Figure 8.1. Map of countries in the WHO Programme.

defaults are actively followed up. National centers should report at a minimum monthly frequency, with preliminary reports if full details and evaluations are incomplete. The data are, however, heterogeneous and subject to all kinds of influences, and the WHO program has agreed on the following caveat to be used by all who produce analyses based on the data:

interpretations of adverse reactions data, and particularly those based on comparisons between pharmaceutical products, may be misleading. The information tabulated in the accompanying printouts is not homogeneous with respect to the sources of information or the likelihood that the pharmaceutical product caused the suspected adverse reaction. Some describe such information as "raw data". Any use of this information must take into account at least the above.

Spontaneous reporting systems are still most frequently used for the detection of new drug safety signals. In developing countries they are also used for the detection of substandard and counterfeit drugs. In all countries about half the adverse reactions that take patients to hospitals have been judged to be avoidable. Now, therefore, there is also a need to consider medical error as a signal that there is a problem with a medical product.

A new trend in spontaneous reporting is the growing number of reports directly from consumers rather than

health professionals. Just as health professionals' reports tell of their concerns about drugs, so do consumer reports and from a different and important perspective.

In all these dimensions, spontaneous reports generate data about possible ADRs, or about broader problems with drugs; they provide the basis for further analysis or for hypotheses for systematic studies. The next steps are:

- to prove or refute these hypotheses;
- to estimate the incidence, relative risk, and excess risk of the ADRs;
- to explore the mechanisms involved;
- to identify special risk groups.

In some unusual circumstances, spontaneous reporting can be used to provide valuable information for these latter tasks as well, but each case requires its own analysis and the exercise of expert clinical judgment.

A recent WHO publication (2002) highlights the new challenges for pharmacovigilance:

Within the last decade, there has been a growing awareness that the scope of pharmacovigilance should be extended beyond the strict confines of detecting new signals of safety concerns. Globalization, consumerism, the explosion in free trade and communication across borders, and increasing use of the Internet have resulted in a change in access to all medicinal products and information on them.

These changes have given rise to new kinds of safety concerns, such as:

- illegal sale of medicines and drugs of abuse over the Internet;
- increasing self-medication practices;
- irrational and potentially unsafe donation practices;
- widespread manufacture and sale of counterfeit and substandard medicines;
- increasing use of traditional medicines outside the confines of the traditional culture of use;
- increasing use of traditional and herbal medicines alongside other medicines with potential for adverse interactions.

According to the same publication, the specific aims of pharmacovigilance are to:

- improve patient care and safety in relation to the use of medicines and all medical and paramedical interventions;
- improve public health and safety in relation to the use of medicines;
- contribute to the assessment of benefit, harm, effectiveness, and risk of medicines, encouraging their safe, rational, and more effective (including cost-effective) use;
- promote understanding, education, and clinical training in pharmacovigilance and its effective communication to the public.

As pharmacovigilance has evolved, the scope of the WHO Collaborating Centre has been extended accordingly, as reflected in the Centre's new vision and goals, and the introduction in the mid-1990s of a new working name, the Uppsala Monitoring Centre (UMC).

DESCRIPTION

OVERVIEW OF ADR REPORTING SCHEMES

ADR reporting schemes differ in a number of dimensions. There are two parallel more or less global systems.

1. The medical literature: many journals publish case reports of patients who experience possible ADRs.
2. National pharmacovigilance systems: case reports of suspected ADRs are collected by national pharmacovigilance centers.

The focus of this chapter is on the second of these (though case reports from the literature are included in some of the

national systems). In this second category there are now two international systems:

1. One under the auspices of WHO in which data on all suspected ADRs are pooled and coordinated by the UMC in Sweden.
2. The European Union (EU) pharmacovigilance system.

In the latter, all Member States and the European Medicines Agency (EMA) are connected via secure intranet (Eudranet) for the exchange of pharmacovigilance information. A database, Eudrawatch, is under development for the collation and analysis of reports of serious ADRs associated with products authorized through the EU centralized procedure. It should be noted that all European countries also belong to, and report to, the WHO system and that the EMA has access to the WHO database information.

Information on various national pharmacovigilance centers has been compiled and in the future this will be updated on the UMC's website (www.who-umc.org). National systems themselves are organized in many different ways. Most are centralized, but an increasing number are decentralized. For most of the national systems the reporting of ADRs is voluntary, but for some it is mandatory. Most national systems receive their reports directly from health practitioners. Some, however, receive most of their reports from health practitioners via pharmaceutical manufacturers, including the largest national system, that of the US (see Chapter 7).

Most centers review each report on an individual basis using a clinical diagnostic and decision-making approach, making judgments about each case as to how likely it is that the drug caused the adverse event (see also Chapter 17). However, others use mainly an aggregate or epidemiological approach to the analysis of the reports (see Chapter 7). Finally, the national centers differ dramatically in how they interact with reporters of ADRs. Some treat their reporters anonymously, providing feedback only in the form of regulatory actions or occasional published papers. Others provide very direct feedback—verbal, written, and/or published—to maximize the dialog between the reporters and the center. Guidelines are available on setting up and running a pharmacovigilance center (UMC, 2000).

ORGANIZATION, AFFILIATION, AND TASKS OF NATIONAL MONITORING CENTERS

In most countries, the monitoring center is part of the drug regulatory authority.

- In some, The Philippines and New Zealand, for example, the functions are carried out jointly by the drug control authority and a university institution. The latter receive the initial reports and perform analyses for consideration by the regulatory authority.
- In Germany, the ADR monitoring program was originally organized by the Drug Commission of the German Medical Profession. In 1978 the responsibility for the evaluation of drug-induced risks was transferred to the National Institute of Health (Bundesgesundheitsamt), and in 1993 a new agency was formed for control of medicines and devices (BfArM). The Drug Commission still collects and evaluates ADR reports from physicians and pharmacists, which then are relayed to the health authorities.
- In France, the French Medicines Agency has taken up duties formerly carried out by the Ministry of Health. It also serves as a coordinating and executive body for a network of 31 regional centers that are connected to major regional university hospitals. Each center is responsible for ADR monitoring in its region. The evaluated reports are fed into a central database. The regional centers are co-sponsored by the agency, the hospitals, and the universities. Other public or even private sources of support can be used as well, provided they are ethical, receive reports, and are authorized by the agency.
- Argentina, Canada, Spain, Sweden, and Thailand also have developed decentralized systems, in parts similar to that of the French.
- In the United Kingdom there are four selected regional centers connected to university hospitals, which have a special responsibility for stimulating ADR reporting in their particular areas.

Regional systems have the advantage that good communication and personal relationships may be established between the staff of the monitoring center and the reporting professionals. They are, however, demanding in the number of staff needed and, unless the reports are fed directly into a central database, can result in delays in the flow of information.

In Morocco, Tanzania, and some other countries the national centers also function as Poison Information Centers, or are closely related to drug information services. These may serve as useful models for other countries, since intoxication and adverse reactions are often related. Also, requests for drug information are often about adverse drug reactions, which may be well known or rare and unexpected. This may further add to the value of a center, as the local physicians then feel that they not only feed in reports of ADRs, but in return receive clinically useful information.

Some regional centers, e.g., in Barcelona (Spain), Bordeaux (France), and those in Sweden, are also engaged in formal pharmacoepidemiology studies to follow up potential signals created by the spontaneous reports.

In many countries ADR monitoring starts within a hospital or academic setting, and may also continue that way. The original activities of the Boston Collaborative Drug Surveillance were a prime example of such an approach. Several other countries, including India and Thailand, as well as some mentioned below, have strong individual hospital monitoring. (See Chapter 27 for more on hospital pharmacoepidemiology.)

REPORTING REQUIREMENTS

The greatest need for information on undesirable and unexpected drug effects relates to drugs that are newly marketed. Thus, most countries emphasize the need to report even trivial reactions to new drugs, while for established medicines only serious reactions are usually requested.

Some countries have clearly identified which new drugs they want observed most closely. In the United Kingdom such drugs are marked with a black triangle in the British National Formulary. The Marketing Authorization Holders (MAHs) are encouraged to include it in all other product information and advertisements. This system is voluntary and in particular cannot be enforced for centralized products (i.e., those approved by the EMEA for use in the European Union).

In Denmark and Sweden, a list of drugs of special interest for monitoring is published in the national medical journal. In New Zealand and Ireland, some selected new drugs are put in an intensive reporting program. 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 ADRs. Most countries, however, issue rather general recommendations as to what type of reactions should be reported to national centers.

In at least ten countries, it is mandatory for physicians and dentists to report cases of suspected serious adverse reactions to the regulatory authority. Both the Council for International Organizations of Medical Sciences (CIOMS) and the International Conference on Harmonisation (ICH) have worked on good pharmacovigilance practices, which set guidelines for the proper management of individual cases and case series. (See also www.ich.org, safety topics.) The US is working on such guidelines as well (see Chapters 6 and 7).

In some 25 countries, including the EU, Japan, and the US, it is obligatory for pharmaceutical companies (or MAHs in the EU) to submit to the regulatory authority cases of suspected adverse reactions that have become known to them (clinical events that might have been caused by the drug, and sometimes those where no such attribution has been made).

SOURCES OF REPORTS

The regulatory status and organization of a national drug monitoring program also determines the sources and the type of information that will be received. Three main groups of countries can be identified:

- Countries obtaining a substantial contribution of reports directly from physicians in hospitals and general practice, such as Australia, France, Ireland, the Netherlands, New Zealand, the nordic countries, Spain, Thailand, and the United Kingdom.
- Countries receiving a vast majority of their information via the pharmaceutical industry, such as Germany, Italy, and the US.
- Countries mainly dependent on information from hospital physicians only, such as Japan, India, Romania, and Bulgaria.

The contribution from dentists is generally small. Some countries accept reports from pharmacists, nurses, and consumers.

HANDLING AND EVALUATION OF REPORTS

When a report reaches a national center a physician or a pharmacist normally reads it. (In some countries pharmacists at the national centers have access to medical consultants.) A judgment is made about whether the information provided is sufficient as the basis for an opinion on the correctness of the diagnosis and causality, or if more data should be requested. In a majority of countries participating in the WHO scheme, the medical officer makes an assessment of each case with regard to the probability of a causal relationship between the clinical event and the drug(s) administered. In many countries an advisory committee of clinical experts helps the national center in making the final causality assessment and the evaluation of the clinical importance of the cumulative reports. The WHO has a system of classifying the summary reports it holds according to their content: a "Quality Grading" based on a publication by Edwards *et al.* (1990).

In recent years there has been an international effort to harmonize the terms used to describe the adverse events and to set criteria and definitions for at least the major serious types of reactions. Similarly, there have been efforts to harmonize the way data are stored and communicated internationally. The main agencies involved in this work have been the WHO, CIOMS, ICH, and the EU. For example, internationally agreed criteria and definitions have been published for reactions frequently reported to the WHO database and by some other groups involved with ADR monitoring. The Medical Dictionary for Regulatory Activities (MedDRA) is becoming more and more used throughout the world, and the ICH E2B format, which is a guideline for the transmission format for information to be included on an adverse reaction case report, and the corresponding IT message specification for transmission, ICH ICSR DTD (Individual Case Safety Reports Document Type Definition), are on the way to being the global data storage and transfer standards for the world.

No common standard for the detailed operational assessment of causal relationship between drug and ADR has been agreed internationally. Most experts agree about which factors should be taken into account in the assessment, but how much weight should be given to each of the factors is the subject of continuing scientific debate (see Chapter 17). A number of more or less complicated and comprehensive algorithms for the assessment of causality have been constructed. When tested by their inventors, these algorithms have, in general, been found to decrease variability among ratings produced by different individuals. This has not, however, always been the case when independent groups have tested the algorithms. Moreover, it has not been possible to test whether the assessments reached by the use of algorithms have been more valid than those reached without them. No algorithm has yet been constructed that can cope with the wide varieties of exposure-event categories seen by a national center and yet is simple enough to be used when evaluating a large number of cases on a routine basis.

The only country today using an algorithm on a routine basis for the assessment of causality in ADR reports is France, where the existence of 31 different regional centers necessitates some standardization. Some national centers are of the opinion that causality rating of each single case as submitted introduces bias, and that it is an unacceptable allocation of resources. It is therefore better to use the term "relationship" since this does not imply a value judgment. However, an international agreement has recently been reached among the countries participating in the WHO drug monitoring scheme on common definitions of the terms most often used to describe relationships in a

Table 8.1. Terminology for causality assessment

Certain. A clinical event, including laboratory test abnormality, occurring in a plausible time relationship to drug administration, and which cannot be explained by concurrent disease or other drugs or chemicals. The response to withdrawal of the drug (de-challenge) should be clinically plausible. The event must be definitive pharmacologically or phenomenologically, using a satisfactory re-challenge procedure if necessary.

Probable/likely. A clinical event, including laboratory test abnormality, with reasonable time sequence to administration of the drug, unlikely to be attributed to concurrent disease or other drugs or chemicals, and which follows a clinically reasonable response on withdrawal (de-challenge). Re-challenge information is not required to fulfill this definition.

Possible. A clinical event, including laboratory test abnormality, with a reasonable time sequence to administration of the drug, but which could also be explained by concurrent disease or other drugs or chemicals. Information on drug withdrawal may be lacking or unclear.

Unlikely. A clinical event, including laboratory test abnormality, with a temporal relationship to drug administration which makes a causal relationship improbable, and in which other drugs, chemicals, or underlying disease provide plausible explanations.

Conditional/unclassified. A clinical event, including laboratory test abnormality, reported as an adverse reaction, about which more data are essential for a proper assessment or the additional data are under examination.

Unassessable/unclassifiable. A report suggesting an adverse reaction that cannot be judged because information is insufficient or contradictory, and which cannot be supplemented or verified.

semi-quantitative way (Table 8.1). Methods for assessing relationships in case reports are discussed in more detail in Chapter 17.

An apparent causal relationship in a single case, or even a series, is not the only issue in comprehensive early signal detection. Many case reports with limited information might be excluded from serious consideration, but a case record that does not allow for remote assessment of the relationship between drug and ADR does not mean that the original observer was incorrect, only that the observation cannot be confirmed. Thus quantity, as well as quality, of reports of associations is valuable. The use of "poor quality" reports as a trigger for a signal should be taken seriously if the clinical event is serious. Early warning is the goal, and a signal based on doubtful evidence should promote the search for better evidence.

There also may be certain items of information within a set of reports that trigger consideration of a signal other than

just the medicinal product and clinical event: the apparent over-representation of higher doses of the relevant drug, or concomitant treatment, or certain patient characteristics may be of interest for the safe use of the drug product.

The above are just some of the common issues for consideration during the evaluation of an early signal. There are many others, such as the finding of a problem with one medicinal product, which triggers a search into products with similar effects. What is clear is that there are very complex interacting patterns of information, which may trigger ideas and concerns.

Many national regulatory authorities systematically review and evaluate information from a variety of sources, in addition to spontaneous ADR reports, to identify new ADRs or changing ADR profiles on the basis of which action should be initiated to improve the safe use of medicines. The web and review journals such as *Reactions Weekly* (ADIS International) are useful in this respect. (*Reactions Weekly* also links its literature findings to those of the WHO database where relevant.)

FEEDBACK TO REPORTERS

Some form of feedback from the national center must be arranged for clinicians to feel that they are involved in an iterative and progressive process. In many countries, each reporter receives a personal acknowledgment, often including a preliminary evaluation of the case. Adverse reaction bulletins are produced regularly in many countries and then distributed to the medical profession. Sometimes the information is included in a local medical journal or a drug information bulletin.

This is, perhaps, the central point at which effective communications are essential for the success of pharmacovigilance. Once physicians know of their national reporting system, believe reporting is important, know where to find their reporting form, and feel motivated to act (all major communications and motivational challenges in themselves), they must feel that their efforts have some reward (recognition, at least) and some effect on medical knowledge and practice.

DETECTION AND EVALUATION OF SIGNALS

Spontaneous adverse drug reaction reporting is principally a method of identifying the previously unrecognized hazards of marketed medicines. Within the WHO program a "signal" concerns information regarding a possible relationship between a drug and an adverse event or drug interaction. In trying to detect signals from international data it should be understood that a signal is an early hypothesis,

and that it simply calls for further work to be performed on that hypothesis.

In the early days of pharmacovigilance, when reports were relatively few, signals were looked for manually or through checking, for example, quarterly lists of submitted case reports sorted in various ways to help review (e.g., all deaths, new-to-the-system). Profiles based on the proportion of reports regarding different system organ classes were compared and differences in the proportion of reactions reported were used as prompts for further analyses. Later, differences in such proportions were tested by statistical significance tests. A published signal, for example, based on such a test was the higher proportion of serum sickness-like reactions to cefaclor, in comparison with other cephalosporins and ampicillin.

The French “case–non case” method is based on the same principle, comparing the proportion of, for example, hypoglycemia reported for acetylcholinesterase (ACE) inhibitors with that reported for other cardiovascular drugs.

The human brain is excellent at finding significant patterns in data: humans would not have survived if that were not so! It is a complex process to examine large numbers of case reports for new factors that may impact upon the safe use of the drug or drugs concerned, especially when for each case report there is a considerable amount of information. Being able to remember the adverse reaction terms used on different reports, how they might be inter-related, and their time trend of reporting is just a hint of such complexity. The vast volume of data in drug safety today cannot be given effective attention, let alone held in the human memory for analysis. Although many important signals have had their origin in open-minded clinical review of data, some presorting is now necessary for the reasons stated above. Also, in order not to miss possible important signals there is a place for subjecting the data to analysis in ways that allow us to see patterns without our preconceptions blinding us to possibilities outside our conditioned experience.

It is true that in looking for significant patterns by sifting through data, something which *looks* probable may turn up by chance: data “dredging” or “trawling” or a “fishing expedition” is bound to catch something, but not necessarily much that is useful. Data dredging should be used as a pejorative term for unstructured fiddling about with data, or worse, the application of a structure to data to make it fit a biased hypothesis in a way to give added credibility to the result.

Formal *data mining*, or “knowledge finding/detection,” on the other hand, is not a random rummaging through data in an aimless fashion, which is what the term “dredging”

implies. Data mining/knowledge finding should be considered as a term for the application of a tool or tools to analyze large amounts of data in a transparent and unbiased fashion, with the aim of highlighting information worth closer consideration.

It is certainly true that the involvement of the variables and the characterization of any of their relationships in advanced pattern recognition is “unsupervised” in data mining, in that a predetermined logic for finding patterns is allowed to run free of human interference. However, in signaling methods using data mining, the level of flexibility and the kind of logic that are applied to data are systematic and transparent.

Data mining approaches to signal detection may use different methodologies but they have in common that they look for “disproportionalities” in data, i.e., a relationship or pattern standing out from the database background experience.

Two of the advantages of these approaches are:

- no external data are needed and the limitations of such data (including delay in receipt) do not apply;
- they may be expected to counteract some of the biases related to variable reporting. For example, if the overall level of reporting is high because of new drug bias, this will not necessarily affect the proportion of all reactions for the drug; while the high overall reporting rate can be related to the extent of drug use, a specific adverse reaction may still be disproportionally reported if it is common.

One possible disadvantage is that, as the background data changes for all drugs in the data set, so does the expectedness (disproportionality) for the drug–ADR combination in question. Stratification will also have the same effect by altering the background data included. This can, however, be taken into account during analysis. It is clear that in a large database the addition of new information to the background will have relatively less influence.

Two approaches to knowledge finding are described below, in some detail, as examples: one for identifying complex patterns in data, and the other for looking at relatively simple relationships.

A Bayesian Approach Using Neural Networks as Implemented for the WHO Database Description

The main use of the WHO program’s international database is to find novel drug safety signals: new information. One begins to see the problem as looking for the proverbial

“needle in a haystack.” As noted above, if important signals are not to be missed, the first analysis of information should be free from preconception.

With this in mind, the UMC has developed a signal detection system that combines a data mining tool for screening of raw data with the subsequent application of different filtering algorithms. This quantitative filtering of the data is intended to focus clinical review on the most potentially important drug–ADR combinations, which can be likened to a clinical triage system for guiding clinical review towards areas of first concern. The resulting “priority package” is scrutinized by independent experts on an international review panel. Based on evaluations of the available evidence and expert opinion, hypotheses of potential drug-induced problems are formulated as “signals.” These are circulated to all national centers participating in the WHO program for consideration of public health implications.

The first step towards this new signaling process was the development of a data mining tool for the WHO database. The Bayesian Confidence Propagation Neural Network (BCPNN) methodology (Bate *et al.*, 2001) was designed to identify statistically significant disproportionalities in a large data set, with a high performance, to allow for automated screening of all combinations of drugs and adverse reactions in the WHO database (Figure 8.2).

The BCPNN provides an efficient computational model for the analysis of large amounts of data and combinations of variables, whether real, discrete, or binary. It is robust

and relevant results can still be generated despite missing data. The missing data do not prevent the identification of disproportionally reported drug–ADR or other combinations; only the uncertainty is greater and denoted by wider confidence limits. If required, it is also possible to impute values, and to create best-case, worst-case information. This is advantageous as most reports in the database contain some empty fields. The results are reproducible, making validation and checking simple. The BCPNN is easy to train; it only takes one pass across the data, which makes it highly time-efficient. Only a small proportion of all possible drug–adverse reaction combinations are actually non-zero in the database. Thus, use of a sparse matrix method makes searches through the database quick and efficient.

The BCPNN is a neural network where learning and inference are done using the principles of Bayes’ law. Bayesian statistics fit intuitively into the framework of a neural network approach as both build on the concept of adapting on the basis of new data.

The new signaling system uses the BCPNN to scan incoming ADR reports and to compare them statistically with what is already stored in the database, before clinical review. Every three months, the complete WHO database is scanned to produce the combinations database. This is a table that contains frequency counts for each drug, for each ADR, and for each drug–ADR combination for which the UMC has received case reports during the last quarter. Only combinations where the drug has been reported as

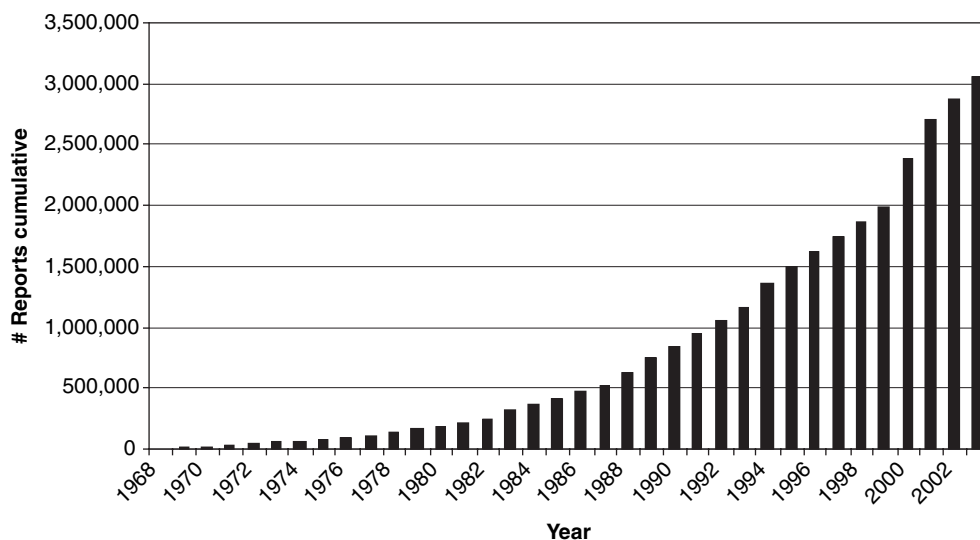


Figure 8.2. Cumulative number of reports in the WHO database 1968–2003.

“suspected” are included. For each drug–adverse reaction combination, statistical figures generated by the BCPNN are also given. The figures from the previous quarter are also included and the data are provided to all national pharmacovigilance centers in an electronic format.

The neural network architecture allows the same framework to be used both for data mining/data analysis as well as for pattern recognition and classification. Pattern recognition by the BCPNN does not depend upon any *a priori* hypothesis, as an unsupervised search and detection approach is used. For the regular routine output the BCPNN is used as a one-layer model, although it has been extended to a multilayer network. To find complex dependencies that have not necessarily been considered before, a recurrent network is used for investigations of combinations of several variables in the WHO database. Two important applications based on the BCPNN that the UMC is developing are syndrome detection and identification of possible drug interactions. Other possibilities include finding age profiles of drug–adverse reactions, determining at-risk groups, and searching for dose–response relationships.

Naturally, changes in patterns such as patient groups, drug classes, organ systems, and drug doses may also be important. However, as with any subdivision of data, a very large overall amount is necessary initially to attain statistical significance in subsets. This is a major advantage of using the large, pooled WHO database, and the UMC is trying to maximize this potential.

Stratification of data has the same problem of needing a large data set as well as the problem of deciding in advance what strata may be relevant in any particular situation, although it may be valuable in removing the effects of confounders. Stratification is, therefore, done after signals have been found, as deemed necessary.

“Validation” of the BCPNN Data Mining Approach

Critics of data mining can reasonably suggest that, with all the possible relationships in a huge database, many medicine–adverse reaction associations will occur by chance, even though they seem to be significantly associated. The BCPNN methodology used by the UMC does take account of the size of the database in assigning probabilities. It is clear that national centers and reviewers must not be provided with what amounts to a huge amount of useless probabilistic information. On the other hand, it is clear that the process of finding signals early will entail some false positives.

Determining the performance of the BCPNN is a difficult task because there is no gold standard for comparison and

there are different definitions of the term signal. According to the definition used in the WHO program, a signal is essentially a hypothesis together with data and arguments, and is not only uncertain but also preliminary in nature: the status of a signal may change substantially over time, as new knowledge is gained.

Two main studies of the performance of the BCPNN have been reported in a single paper (Lindquist *et al.*, 2000).

One study concerned a retrospective test of the BCPNN predictive value in new signal detection as compared with reference literature sources (Martindale’s Extra Pharmacopoeia, and the US Physicians Desk Reference). The BCPNN method detected signals with a positive predictive value of 44% and the negative predictive value was 85%.

The second study was a comparison of the BCPNN with the results of the former signaling procedure, which was based on clinical review of summarized case data. Six out of ten previously identified signals were also identified by the BCPNN. These combinations all showed a substantial subsequent reporting increase. The remaining four drug–ADR combinations that were not identified by the BCPNN had a small, or no, increase in the number of reports, and were not listed in the reference sources seven years after they had been circulated as signals.

Of course, the use of the selected literature sources as a gold standard is open to debate. The literature is not intended as an early signaling system, and uses many sources for its information other than the WHO database: the biases affecting inclusion and exclusion of ADR information therefore may be very different. Factors such as those affecting the differential reporting to WHO and the inclusion of new information in the reference sources will have an effect which is independent of the performance of the BCPNN.

The BCPNN is run every quarter by the UMC, and just one quarter was selected: since the BCPNN is used in continuous analysis, the specificity and sensitivity are subject to necessary time-dependent changes in classification of “positives” and “negatives.” It is difficult to consider something as a “non-association” because of this time dependency, and it is clear that there is an asymmetry in the effect of time on the results.

An assumption was made that a substantial increase in the number of reports of an association over the period indicated ongoing clinical interest in an association. More reports may be seen as a support for the validity of the associations, though there is often a tendency for ADRs that are becoming well known to be more reported anyway.

Another obvious limitation is that the BCPNN method for signal detection is dependent on the terminology used for recording of adverse reactions. Very little work has been

done on any of the medical terminologies in use or proposed to determine their relative value in searching for new drug signals.

Although the UMC found that the use of the BCPNN gave a 44% positive predictive value, and a high negative predictive value of 85%, the usual approaches for assessing the power of a method are difficult to apply, because of the reasons outlined above. Further, and importantly, negative predictive value will always be high when the *a priori* probability is low. Thus, an 85% negative predictive value may not even be perceived as high in this situation. It is for these reasons that “validation” is placed in quotation marks in the title of this section.

The BCPNN (or indeed any other data mining method) is not a panacea for drug safety monitoring. It is important to be aware of the limitations of the BCPNN and that it cannot replace expert review. However, it may be a very useful tool in preliminary analysis of complex and large databases.

Proportional Reporting Ratios

This is an approach pioneered in the UK but now widely used because of its simplicity. The proportion of all reactions to a drug that represent a particular medical condition of interest is compared with the same proportion for all drugs in the database. The resulting statistic is called a proportional reporting ratio (PRR). Judgments about signals may then be made using the PRR, along with the associated value of chi-squares and the absolute number of reports.

As with the other methods, this approach uses the total number of reports for the drug as a denominator to calculate the proportion of all reactions that are the type of interest (e.g., hepatitis). This proportion may be compared with the value for other drugs. It is also possible to compare complete profiles of ADR reporting for drugs of different types of reactions, where differences in the profile may represent potential signals. The result of such a calculation is also called a PRR, where the PRR is $a/(a+b)$ divided by $c/(c+d)$ in the following two-by-two table:

	Reaction(s) of interest	All other reactions
Drug of interest	<i>a</i>	<i>b</i>
All other drugs in database	<i>c</i>	<i>d</i>

The result of the PRR, as well as the other methods described below, is to highlight drug–ADR combinations

with a disproportionally high reporting rate: the mathematics involved are different but the principles are similar.

The expected or null value for a PRR is 1.0 and the numbers generated are measures of association that behave in a fashion similar to relative risks. Measures of statistical association for each value are calculated using standard methods for significance (see below). The higher the PRR, the more the disproportionality stands out statistically. Examination of changes in PRRs over time may help to demonstrate how disproportionalities can be identified as early as possible.

PRRs have advantages over calculation of reporting rates, since they are simple in concept and calculation, and enjoy the merits of data mining in general. However, and importantly, this does not take into account clinical relevance and the effects of biases including confounding, selective under-reporting, and the effects of summarizing report information (see Case Example 8.1).

CASE EXAMPLE 8.1: DATA MINING APPROACH TO CASE REPORTS

Background

- Case reports from health professionals on suspected adverse outcomes from drug therapy provide the basis for hypotheses about new drug adverse reaction relationships, patient safety issues such as drug interactions, dose–effect relationships, and much other information about problems and uncertainties which health professionals and patients experience in the use of drugs.

Issue

- The difficulty is that there are huge numbers of case reports sent to regulators and companies, of very variable quality.
- The number of potentially useful data fields is also large, including background information about the patients, their diseases, all the drugs they take, doses, routes of administration, and more.
- The numerator of the volume of reports needs to be assessed against some background of drug use to gain some first impression of the importance of the drug–suspected-adverse-reaction relationship compared to similar drugs and the extent of their use.

Approach

- It is always useful to know the extent of use of the drug, but the information is often either not available,

or only in the crudest terms such as the numbers of tablets or even the tonnage of drug sold.

- Given that good drug utilization information may take some time to find, and that only crude quantification is possible (and needed) from case reports, another option is to compare drugs, one to another or by various groupings, by disproportionality of the ratio:

$$\frac{(\text{Target adverse reaction number} \div \text{total number of other reactions for target drug})}{(\text{Target adverse reaction number} \div \text{total number of other reactions for comparator drug/drugs})}$$

- Doing this utilizes “all other reaction” numbers as a kind of surrogate for comparative patient exposure to the drug, as well as a comparative profile of possible adverse drug reactions.
- It is also possible to use a variety of other techniques, such as unsupervised pattern recognition, multivariate analysis, and stratification to look for patterns in the data from case reports.
- In view of the large quantities of data that might be analyzed, data mining/knowledge detection approaches are often used on neural network IT platforms, though the basic approaches are very simple.

Results

- Analysis of case report information finds suspected new adverse reactions, with variable certainty. One study on a large amount of data showed a positive predictive value of nearly 50%. The assessment of predictive value is very uncertain, however, because of the lack of a “gold standard”.
- Patterns in case data indicate the areas and parameters for further more definitive study.

Strengths

- Case reporting is the only way to determine a range of health professional and patient concerns about drugs.
- It is relatively inexpensive, universal, and continuous in its coverage.
- It is non-interventional.
- Data mining allows huge amounts of information to be screened in an objective way for novel patterns within the data.

- Data mining allows for some analysis of biases and influences on reporting rates.

Limitations

- Case report information is subject to many and variable outside influences, which introduce biases as well as there being variation in the extent of underreporting.
- That data mining includes numbers and statistics can confer an erroneous certainty on the results.
- Expert scrutiny of results is essential, and hypotheses should be evaluated further using other methods.

Summary Points

- Case information allows health professionals and patients to share adverse experiences with drug treatment, sometimes in great detail.
- Data mining techniques allow complex searches in large volumes of case data to find relationships and patterns that stand out, and therefore deserve further scrutiny.
- Statistical values in data mining reflect the robustness of relationships and the distinction from background data. They do not indicate the nature of the relationships, which easily may be due to biases.
- The results of data mining should not be confused with nor interpreted as if they were the results of formal, controlled studies.
- Whether data mining is used or not, expert human examination of case data is essential to determine what further action is demanded and useful.

It is again important to recognize that PRRs are not a substitute for detailed review of cases, but an aid to deciding which series of cases most warrant further review. Also, PRRs and chi-squared values are measures of association and not causality. The result of a PRR provides a signal; it does not prove causation. Testing of the resulting hypothesis usually requires formal study in more structured data.

There are a number of possible extensions to the method that are being evaluated further, for example by Professor Stephen Evans in the UK, using the differences in observed and expected reporting rates (sequential probability ratio test) as a way of highlighting the more important possible signals (personal communication). Also, PRR calculations could be restricted to particular groups of drugs, to serious or fatal reports, or to particular age groups.

Other Data Mining Approaches for Signal Detection

Data mining approaches have been adopted by several national pharmacovigilance centers, including the Netherlands, the UK, and the US and also by some pharmaceutical companies. A recent paper (Evans *et al.*, 2001) analyzed the concordance among different measures (PRR, reporting odds ratios (ROR), and the *IC*). The different methods all use the principle of finding disproportionality as described for the PRR, but the mathematical theory behind them is different, for example, a Bayesian approach and information theory underpinning the BCPNN. The result of the comparative study was that no clear differences were present, except for when there were fewer than four reports per drug–ADR combination, though this depends on the Bayesian prior probability used in those methods using Bayes' theory. The methods all have their somewhat different advantages and disadvantages. Before a partially automated signal detection system is implemented, it is therefore recommended that careful consideration be given to the possible alternatives in looking for the most practical and appropriate in a given setting.

QUANTITATIVE ASPECTS OF REPORTING

Case reports are submitted by national centers to the UMC for inclusion in the WHO database. Figure 8.2 depicts the cumulative number of reports stored in the database. In most countries reporting has gradually increased over time. It usually takes some 5–10 years of operation before reporting reaches a stable level. The number of reports relayed to the UMC is often less than that received in the country, for various technical reasons. In France, reports received by the national agency from the manufacturers (50%) are not sent to the UMC, nor are reports on drugs that are marketed only in France. Also, reports evaluated as unclassifiable or reactions due to overdoses are omitted from some but not all countries. The US situation is special in this regard and is described separately (see Chapter 7). (The issue is also discussed further below; see “Limitations”.)

INTERNATIONAL REPORTING STANDARDS

There is no international standard for good pharmacovigilance practice but some useful publications give advice regarding some aspects (CIOMS, 2001; Meyboom, 1997, 2000).

In recent years there has been an increasing tendency among regulators to demand that manufacturers report suspected ADRs occurring in other countries directly to them, in spite of the fact that most of such reports are already available to

them through the WHO system. At first, these requirements also meant that the manufacturers had to report these cases on several different forms, according to different rules and time schedules.

In order to decrease the workload of manufacturers and increase the cost-effectiveness of international reporting, an initiative was started in 1986 to harmonize the rules for international reporting of single cases under the auspices of CIOMS, which is affiliated to the WHO. This initiative, called “CIOMS I,” is now accepted in most countries and by most manufacturers, and has been accepted by the ICH as a guideline for use in Japan, the European Union, and the US. This system has definitely decreased the diversity of rules in international reporting, although national variation still exists even in the ICH member countries. There is also a CIOMS guideline on electronic reporting of adverse reactions internationally.

Both inside and outside the ICH, many of the smaller drug control agencies felt that they did not have the capacity to cope with the massive increase in the number of reports that the CIOMS I initiative produced. They preferred periodic safety updates, including the evaluation of the safety situation at large by the company. In some countries such safety updates were mandatory, but again there were differences in the rules, formats, and time schedules. Therefore, a second CIOMS project—“CIOMS II”—was initiated to harmonize the contents, format, and time schedule for periodic safety updates. Some novel features of this scheme were:

- the creation of an international “birth date” for each drug product, which was the day of first approval in any country;
- the inclusion of drug exposure data and experience from both pre- and post-marketing studies;
- the principle that the manufacturer should write an overall safety evaluation of the product.

The CIOMS II safety updates are now unofficially accepted by ICH countries and have been made into a guideline followed by the EU, the US, and Japan. Many more countries accept periodic safety updates according to the CIOMS/ICH format, although again there are some variations in demands by regulatory authorities, which frustrates the concept of complete global standardization that the pharmaceutical industry seeks in order to reduce the administrative burden.

The European Parliament adopted a regulation in 1995 creating the EMEA, which is located in London, UK. According to this regulation, any serious suspected adverse reaction that is reported to an MAH by a health professional must be reported to the health authority of the country in which it occurred within 15 days. Health authorities must

also report to the EMEA and to the MAH within 15 days. MAHs are also required to report ADRs occurring outside the EU as well as from the world literature that are both serious and unexpected (known as *unlabeled*: not fully covered in the Summary of Product Characteristics package insert which accompanies the drug product). Similar regulations exist in the US, described in Chapter 6, and in some other countries. This “15-day reporting” certainly has the advantage of getting prompt warnings of new signals of adverse drug reactions, but needs to be considered against the need for follow-up reports and the possible errors and duplications, which seem to have been difficult to avoid so far.

The problems of managing and distributing large numbers of ADR reports within 15 days has led to a commitment from many countries to communicate data electronically. However, standardization of data sets and terminology is required for this to be implemented effectively. The MedDRA and Electronic Standards for the transfer of Regulatory Information were introduced to the ICH (M1 and M2 topics) in 1994. Both are applicable to the pre- and post-marketing phases of the regulatory cycle (see www.ich.org).

MedDRA was based on the terminology developed by the Medicines Control Agency (now the Medicines and Healthcare Products Regulatory Agency; MHRA) in the UK. It provides greater specificity of data entry (more terms) than previously used adverse reaction terminologies and hierarchical data retrieval categories. However, it does not contain specific definitions of the terms to be used. In this regard, the CIOMS initiative is relied upon. Its use for ADR reporting will be mandated in some countries, but other countries, especially small and developing countries, have expressed doubts since their computer resources are limited and their number of reports is small.

STRENGTHS

Spontaneous reporting systems are relatively inexpensive to operate, although their true total costs to the health care system, including the substantial investment by the pharmaceutical industry in their maintenance, are unknown. They remain the only practical means of covering whole populations and all drugs. They allow significant concerns of health professionals and patients to be expressed, pooled, and acted upon to form hypotheses of drug risks. One physician, one pharmacist, and a secretary can usually manage 1000–2000 reports a year, depending on the amount of scrutiny, follow-up, feedback, and other activities that are part of the program. The basic technical equipment needed

Table 8.2. Strengths and limitations of spontaneous reporting systems

Strengths

Inexpensive and simple to operate
Covers all drugs during their whole life cycle
Covers the whole patient population, including special subgroups, such as the elderly
Does not interfere with prescribing habits
Can be used for follow-up studies of patients with severe ADRs, to study mechanisms

Limitations

The amount of clinical information available is often too limited to permit a thorough case evaluation
Underreporting decreases sensitivity and makes the systems sensitive to selective reporting
The reporting rate is seldom stable over time
No direct information on incidence

is also a relatively minor investment. Together with other qualities (Table 8.2), some of which are unique, this makes a spontaneous reporting system one of the essential, basic ingredients in a comprehensive system for the postmarketing surveillance of drug-induced risks.

A spontaneous reporting system has the potential to cover the total patient population. It does not exclude patients treated for other concomitant diseases or those who are using other drug products. Moreover, the surveillance can start as soon as a drug is approved for marketing and has no inherent time limit. Thus, it is potentially the most cost-effective system for the detection of new ADRs that occur rarely, mostly in special subgroups, like the elderly, or in combination with other drugs.

In an early analysis of how important ADRs were first suspected and then verified, Venning (1983) found that 13 out of 18 reactions were first signaled by an anecdotal report made by a physician with an open and critical mind. The fact that these reports were published in medical journals led the author to conclude that spontaneous reporting systems were of little value in the signaling process. However, the majority of these reactions actually were detected before most spontaneous reporting systems were operational. In later analyses of where the first suspicion that a new drug could cause agranulocytosis and Stevens–Johnson syndrome appeared, it was found that, for the vast majority, the first report appeared in the WHO database more than six months before it was published. The opposite situation was found in only a few cases. Some of these signals were published. In most cases, however, a spontaneous reporting system needs to be supplemented by other sources of information, as described in the section on “Particular applications,” below.

LIMITATIONS

Spontaneous reporting systems are mainly intended to produce signals about potential new ADRs. To fulfill this function properly, it must be recognized that a number of false signals will be produced and, therefore, that each signal must be scrutinized and verified before it can be accepted and acted upon. Preferably a signal should be followed up using an analytic epidemiologic study design and one or more of the data resources described in Chapters 9–13. Signals might be of rare reactions, however, so the constraints of any epidemiologic study are an important consideration. Quite often it is not easily possible to create even a case–control study that will answer questions about rare serious reactions. In such instances, it seems reasonable to qualify the result by a statement such as: “spontaneous reports have suggested adverse reaction *x*, but further studies have been unable to confirm this at an incidence more frequent than . . .”

A more serious disadvantage is that not all reactions are reported and that the proportion that is reported in any specific situation is hard to estimate. A basic requirement for the generation of a report is that a physician suspects that a drug may be causing the signs and/or symptoms of his/her patient. This is relatively easy when the inherent pharmacological actions or chemical properties of the drug can predict the reaction. There are also some diseases that are considered as “typically” drug-induced reactions, such as agranulocytosis and severe skin reactions, so that the basic level of suspicion is high. It is, however, very hard to make the mental connection between a drug therapy and a medical event if the event simulates a common, spontaneously occurring disease or other untoward event, which has never previously been described as drug-induced. Some examples of this include the first cases of the oculo-mucocutaneous syndrome, Guillain–Barré syndrome, and changes in the body fat distribution, structural heart valve changes, and visual field defects with a specific pattern. It is also hard to make the mental connection between a drug and a medical event if there is a long time lag between exposure and disease.

Even if the physician suspects the signs and symptoms of his or her patient to be drug-induced, ignorance of the value of ADR reporting or of the reporting rules, and overwork, have been given as reasons for not reporting. Increased information and feedback by national agencies, medical schools, pharmaceutical manufacturers, and professional journals in collaboration could rectify these inadequacies. The health care authorities also have a clear

role here. It is their responsibility to monitor the quality of health care and to build ADR reporting practices in their quality assurance systems, as well as in continuing medical education.

Besides delaying the detection of new ADRs, underreporting creates two other important problems. First, it will underestimate the frequency of the ADR and, thereby, underestimate the importance of the problem. This may not be so serious as long as one recognizes that the reported frequency is likely to be a minimum level of incidence. More important is that underreporting may not be random but selective, which may introduce serious bias.

The effect of selective reporting becomes potentially disastrous if the number of reports of an ADR for different drugs is compared in an uncritical way. There are many possible reasons for apparent differences. The overall rate of reporting has increased over the years, and reporting is often higher during the first years a new drug is on the market (the “Weber effect”). Finally, a drug that is claimed to be very safe may first be tried on patients who do not tolerate the previous drug products (channeling bias). Furthermore, there may be reporting distortions if there are suspicions or rumors circulated about a drug.

Another interesting example of biased reporting is the four-fold difference in reports of hemorrhagic diarrheas in relation to sales of the two penicillin-v products Calciopen® and KØevepenin® in Sweden. An analysis of the situation failed to reveal any differences in the two products. They were produced in the same factory from the same batch and only the form, product name, and MAH differed. There were, however, differences in the use of the products. The older product was used to a larger extent by older physicians, by ear, nose, and throat specialists, and private practitioners, groups who traditionally do not report ADRs. This could not, however, totally explain the apparent difference in ADR rate, as the reporters were seldom the prescribers, but other health care professionals who managed the patients’ illness. The most important explanation was probably that the newer product was more commonly recommended and used in “high reporting” areas. A similar situation arose when minocycline was found to have a higher reporting rate for the lupus syndrome and hepatic injury, both together and separately, than other tetracyclines. This difference in rate is probably due, at least in part, to the much more protracted use of minocycline in the management of young people with acne, as compared with the shorter-term use of the other tetracyclines to treat infection.

PARTICULAR APPLICATIONS

WITHOUT THE ADDITION OF OTHER DATA

It is rarely possible to use spontaneous reports to establish more than a suspicion of a causal relationship between an adverse event and a drug, unless:

1. there is at least one case with a positive re-challenge and some other supportive cases which do not have known confounding drugs or diseases, or
2. there is an event which is phenomenologically very clearly related to single drug exposure, such as anaphylaxis, or
3. there is a cluster of exposed cases reported, the background incidence of the adverse event is close to zero, and there is no confounding.

Even the reappearance of an adverse event when a drug is given again is certainly no proof of causality, unless this is done in tightly controlled circumstances, which may be unethical. Thus, re-exposure to a drug that has caused an ADR is often accidental. In practice, however, one is reassured that there is strong evidence for a causal relationship if there is a cluster of cases with good clinical information, in which the same event has reappeared with repeated exposure at least once in each patient. This is only possible if the medical event in question is of a type that would diminish or disappear after withdrawal of the drug and not reappear spontaneously. Thus, the observation of five cases of aseptic meningitis that reappeared within hours after again taking the antibiotic trimethoprim for urinary tract infections will convince most clinicians that this drug can and did cause such a reaction. For typical "hit and run" effects like thromboembolic diseases and for diseases that can be cyclic, information on re-challenge, however, can be misleading. For example, one unpublished case report involved a young boy who developed agranulocytosis three times in connection with infections treated with ampicillin. It was not until after the fourth time, when agranulocytosis developed *before* ampicillin was given, that his cyclic neutropenia was discovered. Contrast this with a patient who three times developed the Guillain-Barré syndrome after tetanus vaccination. Such unpublished material has influenced policy.

Information on re-challenge is relatively uncommon in most spontaneous reporting systems. Planned re-challenge may be dangerous, is seldom warranted from a clinical point of view, and can be unethical. However, information on a positive re-exposure was available in as many as 13% of 200

consecutive nonfatal cases reported in the nordic countries. In the French pharmacovigilance database, re-exposure is reported in 8.5% of reports, positive in 6%.

An example of comparing rates of a cluster of events with the known background occurred with the cardiovascular drug aprindine. Four cases of agranulocytosis were reported in the first two years the drug was marketed. As the background incidence of agranulocytosis is only five to eight per million inhabitants per year, this made a strong case for a causal relationship. Similarly, there were 25 confirmed cases of motor and sensory neuropathy following meningitis A vaccination in 130 000 children, which was much higher than the likely background incidence.

These examples show some ways in which spontaneous ADR reports can lead to firm conclusions being made, and such reports may be the only information available, particularly during the early marketing of a product. It should not be necessary to emphasize that reports should be of good quality when they are the only evidence used!

WITH THE ADDITION OF DENOMINATOR DATA

Today, most drug regulatory authorities in industrialized countries and pharmaceutical manufacturers collate information that can be used to estimate both the size and the characteristics of the exposed population and the background incidence of diseases. In many countries there are national statistics on drug sales and/or prescribing (see Chapter 27). In many countries information on drug sales and prescribing is confidential, but in the nordic countries this information is published periodically.

IMS Health is a unique source of information on pharmaceutical sales and prescribing habits in a large number of countries (see Chapter 13). The data are collected continuously on nearly all drug products. The prescription data are obtained from rolling panels of prescribers in each country, constituted to reflect the national mix of medical specialists and medical practice type. These data are not without the usual drawbacks of continuous routine data collection, yet their use has provided many important insights into drug safety issues. For example, data from several countries have been combined with ADR information from the WHO database to provide rough incidence estimates of certain drug-induced problems, and perspectives on possible reasons for differences of ADR reporting have also proved useful.

WITH THE ADDITION OF NUMERATOR DATA

If the rate of reporting is known, the numerator can be inferred with some accuracy. From studies using registers of hospital discharge diagnoses, it has been possible to calculate reporting rates for some areas, for ADRs, and for selected periods of time. Considering serious reactions such as blood dyscrasias, thromboembolic disease, and Stevens–Johnson syndrome, between 20% and 40% of the patients discharged with these diagnoses have been found to be reported. By identifying all positive BCG cultures in bacteriology laboratories, it was found that almost 80% of all children who developed an osteitis after BCG vaccination had been reported. However, reporting rates probably cannot be generalized. The magnitude of underreporting is important to know when evaluating the data, but should not be used to correct for underreporting in the calculations since the reporting rate is time-, problem-, drug-, and country-specific.

EXAMPLES OF USING SPONTANEOUS REPORTS TO ESTIMATE RISK

If information from an efficient spontaneous reporting system can be combined with information on drug sales and prescription statistics, it is possible to start to consider reporting rates as a rough estimate of the frequency or incidence rate of an ADR. Such estimates can never reach the accuracy of those derived from clinical trials or formal epidemiologic studies. However, they can serve as a first indicator of the size of a potential problem, and certainly an indicator of health professionals' concerns about a problem. For very rare reactions, they may actually be the only conceivable measure.

With knowledge of the number of defined daily doses (DDDs) sold and the average prescribed daily dose (PDD) (see Chapter 27), it is possible to get a rough estimate of the total person-time of exposure for a particular drug. The number of cases reported per patient “exposure time” might then be a very rough, preliminary guide to minimum incidence. If prescription statistics are available, the number of prescriptions may be a better estimate of drug use among outpatients than the number of treatment-weeks calculated from sales data, especially where drug use is mostly short term and doses and treatment times may vary with patient age and indication.

If the background incidence of a disease is known or can be estimated from other sources, it is sometimes also possible to calculate rough estimates of relative risks and excess risks from spontaneously reported data on ADRs plus sales and prescription statistics. For example, single cases of

aplastic anemia in patients taking acetazolamide (a carbonic anhydrase inhibiting diuretic which is used mainly for the treatment of glaucoma) have been reported since the drug was introduced in the mid-1950s. There are no estimates of the incidence of this reaction, but it was certainly thought to be very rare, probably rarer than aplastic anemia occurring after the use of chloramphenicol. Between 1972 and 1988, 11 cases were reported to have occurred in Sweden. Based on sales and prescription data, it could be estimated that the total exposure time was 195 000 patient-years during the same period of time, yielding a reported incidence of about 1 in 20 000 prescriptions, or 50 per million patient-years. From a population-based case–control study of aplastic anemia in which Sweden participated, it could be estimated that the total yearly incidence of aplastic anemia in the relevant age groups was about 6 per million exposed persons. In the case–control study it was not possible to estimate the relative risk for the association between acetazolamide and aplastic anemia, because there were no exposed controls. However, if the spontaneously reported incidence of aplastic anemia among persons exposed to acetazolamide is compared with the total incidence of aplastic anemia from the case–control study, the relative risk could be estimated to be around 10.

Several potential sources of errors in this study must be considered. The degree of underreporting in this example is unknown. However, in one study the reporting rate for aplastic anemia was found to be 30%, and since then reporting in general has doubled. There is no known association between glaucoma and aplastic anemia that could act as a confounder, but some of the reported patients had taken other drugs during the six months before the detection of their aplastic anemia. There were only two patients who had been treated with drugs that, on clinical pharmacological grounds, seemed to be reasonable alternative possibilities as causal agents. However, it is a clear limitation that multiple drug exposures cannot be corrected for in a rough analysis such as this.

In some instances it has been possible to compare risk estimates from a formal epidemiologic case–control study with those derived from the Swedish drug monitoring system. The relative risks for agranulocytosis induced by cotrimoxazole and sulphasalazine were astonishingly alike.

USING SPONTANEOUS REPORTING DATA TO IDENTIFY MECHANISMS AND RISK GROUPS

As soon as it has been established that a drug can induce a certain adverse reaction, it becomes important to identify the mechanisms involved, and whether any group of patients

is at a particularly increased risk, and if any measure could be taken at the patient and/or the population level to reduce the risk. Usually a multitude of different methods must be applied, both in the laboratory and at a population level. A good spontaneous reporting system can be of value in this work in certain circumstances, if the data can be compared to sales and prescription data or if the patients can be subjected to special investigations.

For example, in one study of the characteristics of patients developing hypoglycemia during treatment with glibenclamide (an oral antidiabetic drug), the distribution of prescribed daily doses was similar in patients with episodes of severe hypoglycemia and in the general population. However, patients hospitalized because of severe hypoglycemia were older and were more likely to have had a previous episode of cerebrovascular disease.

In studies published on oral contraceptives and thromboembolic disease, it was found that women who were reported to have developed deep vein thrombosis while taking oral contraceptives were deficient in Leiden factor V more often than would have been expected from the general population.

Another study investigated patients reported to have developed lupoid reactions while taking hydralazine for hypertension. A much higher percentage of these patients were slow acetylators than the 40% expected from the distribution of this phenotype in the population at large.

In a more sophisticated study, Strom *et al.* (1989) used spontaneously reported cases of suprofen-induced "acute flank pain syndrome" in a case-control study designed to identify patient-specific risk factors for the development of the syndrome. Patients who were reported to have developed the syndrome were compared to a random sample of patients who had taken the drug without problems. Risk factors identified were, among others, male sex, hay fever and asthma, participation in exercise, and alcohol consumption. Most of these factors are consistent with the postulated pathogenic mechanism of acute diffuse crystallization of uric acid in the renal tubules.

CONTEMPORARY ISSUES: AN EXAMPLE

The monograph CIOMS IV (1998) entitled *Benefit-Risk Balance for Marketed Drugs* (which should really have been *Effectiveness and Risk Balance!*) promotes the idea that the positive side of drug action and the negative side can both be reduced to similar terms to allow comparison between therapies for the same indication. At the first signal of a problem with a drug the first question should be, "How does this drug compare with others?" Most often there is no information on the relative effectiveness of drugs in real-life

practice, only some premarketing efficacy data, in highly selected patients. For this reason, comparisons at the signal stage are often for the safety side only.

Another major thrust of CIOMS IV was that all similar drugs reasonably used for a particular target indication should be compared and that analysis should consider *the whole safety profile* of the drugs, not just a target adverse reaction. Because the safety profiles of drugs are usually made of a multiplicity of adverse reaction terms, CIOMS IV suggested reducing the comparison to a few: some of those most frequently reported ADRs and the most serious, the latter being assessed using clinical judgment and irrespective of reporting rate.

Following a signal that the lipid lowering statin, cerivastatin, was associated with rhabdomyolysis, the UMC carried out a preliminary assessment in 2000. The CIOMS logic was used in this study, but no data on effectiveness were considered. Even given these limitations, the comparison was interesting.

All the statin drugs classified as similar in the WHO Anatomic-therapeutic Chemical Classification (ATC group C10AA) were selected from the WHO database. Gemfibrozil was specifically included as well because it was very strongly implicated as an interacting drug. IMS Health worldwide sales figures for the years 1989–2000 were obtained for the same drugs and from nearly the same countries (IMS does not have data from the Netherlands, Iran, Costa Rica, and Croatia, but these countries contribute a relatively very small number of ADR reports). IMS data are difficult to obtain prior to that date and therefore missed the launch of the first statin on the market by two years. The IMS data were converted to million patient-years and, since the mean dosages used equate closely with dosage forms, these were used as a rate denominator in all subsequent work. The rates are shown in Table 8.3.

Also, IMS Health data on co-prescription between statins and gemfibrozil and fibrates was used, together with age and gender breakdowns of statin prescriptions in the US.

All critical ADR terms (WHO Critical Terms List) associated with the statins were inspected by a clinician, who selected the ADR terms with the most serious clinical import in terms of possible lethal outcome or permanent disability. They were then grouped, resulting in 13 ADR groups:

- disseminated intravascular coagulation
- neuroleptic malignant syndrome
- cardiomyopathy
- anaphylactic shock
- death
- serious hepatic damage

Table 8.3. Sales denominators for statin drugs, 1989–2000

Drug name	Million patient-years
Atorvastatin	22.49
Cerivastatin	3.92
Fluvastatin	8.44
Lovastatin	20.51
Pravastatin	46.55
Simvastatin	44.97

- rhabdomyolysis
- pulmonary fibrosis
- Stevens–Johnson syndrome
- renal failure
- myopathy
- pancreatitis
- serotonin syndrome.

Because of the small numbers, anaphylactic shock, disseminated intravascular coagulation, and neuroleptic malignant syndrome were not evaluated further, and it was considered that the overall report profiles were qualitatively similar.

There were, however, some quantitative differences. Of rates above 1 per million patient-years, the rates that stood out were: cardiomyopathy, myopathy, renal failure, rhabdomyolysis, and death with cerivastatin, and myopathy for lovastatin. This caused the study to focus on the issues of rhabdomyolysis, myopathy, and renal failure as reasonable in comparing and differentiating the safety profiles of the drugs.

The various other analyses that were done can be found in WHO Drug Information (<http://www.who.int/druginformation>). The following is a summary of some of the study's conclusions:

This signal was known for two years before the drug was taken off the market and attempts were made to warn the medical profession of the very important risk of interaction. The result of repeated "Dear Doctor" letters was only a 2% change in prescribing behavior. It was and is clear that the communication of these key messages must be improved and that monitoring impact is an all important practice.

Based on an analysis of spontaneous reports there appears to be a strong link between cerivastatin and rhabdomyolysis and renal failure (possibly related) which was quantitatively greater than with the other statins. Since cerivastatin is the most recently marketed this may increase the overall reporting rate relatively. The Weber effect was not a complete explanation for the rhabdomyolysis reporting.

Disproportionality of the combination cerivastatin/gemfibrozil and cerivastatin/clopidogrel against their use strongly suggests interaction with both drugs. This almost entirely affects muscle

ADRs. This effect of the combination was not seen as strikingly with the other drugs, though it was obvious with lovastatin. Disproportionality of cerivastatin reports of rhabdomyolysis and its use in older women suggests they may be a risk group, though lovastatin was relatively highly used in older women and seemed less likely to cause rhabdomyolysis in this group.

There were clear problems of case definition. The link between myopathy and death on older reports suggests misclassification under the broader, more neutral term. The need for case definition when the ADR involves a rare disease is important and may delay and confuse positive action. It is certainly possible that the profile of lovastatin and muscle disorder and death was not very different to cerivastatin. The availability of rhabdomyolysis as a used term since the mid 90s probably resulted in lower numbers of reports of rhabdomyolysis with older drugs, particularly lovastatin. Reports of myopathy are high for lovastatin after the first years of launch. Lovastatin has been on the market longest and there may be a depletion effect.

The purpose of giving this example study here is not to promote or endorse any of its conclusions, but to indicate the range of issues that are raised. It seems unlikely that the above issues could be answered without very extensive studies, because of the rarity of rhabdomyolysis, but this is an important issue because of the wide and increasing use of the statins. It is just one example of the safety management problems with "blockbuster drugs," and the need for risk management planning. The failure of repeated communications about the interaction between cerivastatin and gemfibrozil raises the need for a much greater attention to communications of risk and their impact. This subject is dealt with elsewhere (see Chapter 27), and has been the topic of two international workshops, both of which resulted in monographs.

THE FUTURE

In western countries, the population is growing progressively older and, thus, we can expect a steady increase in the chronic use of medications. Even if the drugs to be used are more sophisticated and "targeted," they are also likely to be more pharmacologically active and hence may be more difficult to use. With the continued development of clinical trial methodology, adverse reactions that are caused by pharmacologic mechanisms will probably be better known both as to range and incidence when new medicines are approved. However, there will still be the classical idiosyncratic reactions, which cannot be predicted and which are too rare to be detected in the clinical premarketing trials.

At the same time, there is increasing commercial pressure for the pharmaceutical industry to find "blockbuster" drugs

that will be marketed globally to maximize profit in the shortest possible time. Other changes in the industry—shortened times for drug development and increasing outsourcing of functions—make for an environment where some premarketing safety issues may go unnoticed. The increasing challenge to pharmacovigilance is not only to be able to find early signals of drug problems, but to rapidly determine true effectiveness and risk during regular clinical use.

Pharmacovigilance programs are now being established also in developing countries, from which there has been little information in the past. It is likely that the monitoring of populations with other patterns of morbidity and different nutritional status will reveal different types of adverse reactions than what we have learned to expect from populations in the industrialized world, even from established medicines. The influence of co-medication with traditional medicines, and the unexpected failure of efficacy because of substandard or counterfeit medicines, will have to be covered by the pharmacovigilance systems. The developing countries have some diseases that are not seen in the so-called developed world. Malaria is one example only, but this scourge and others are being treated by new chemical entities within large public health programs. Pharmacovigilance and pharmacoepidemiology must be employed alongside such programs if drug-related risk is to be detected early and limited.

The role of spontaneous reporting in the future will be of even more central importance if it can be developed further. The basic requisite for its enhanced effectiveness is an increased flow of information, both in quantitative and qualitative terms. For example, to increase the reporting of classical, rare ADRs such as blood dyscrasias, toxic epidermal necrolysis, and liver and kidney damage, the automatic collection of information about all patients who have been hospitalized with these conditions could be instituted. This could be accomplished through case-control surveillance of rare diseases that are often elicited by drugs. Alternatively, manual retrieval of case summaries, providing high quality information, or alternatively through automated transfer of computerized hospital discharge diagnoses, could be used to study case series. Of course, in such automated systems, one would lose the important screening impact of the provider who thought the adverse outcome was due to the drug. Periodically it may be beneficial to focus on certain new drug classes to clarify their ADR spectrum as soon as possible (the HIV ADR reporting scheme in the UK is an example).

However, the detection of the totally unexpected will most probably continue to rely on the capacity of the alert human mind for the foreseeable future. Therefore, it is essential to enhance the practicing clinician's awareness of and coopera-

tion with ADR reporting. Here a regional system with mutual benefits, like the French system, seems promising.

The WHO and the UMC have taken several initiatives to promote the importance of effective communications. Another publication on crisis management in pharmacovigilance also deals with communications issues, and the publication *Viewpoint* (2002) is aimed at explaining some of the issues involved in pharmacovigilance to the general public. Alongside the development of scientific excellence, while the core of pharmacovigilance is the discovery of knowledge about drug safety, its achievements are of little value if they do not actively affect clinical practice and the wisdom and compliance of patients. The basic aim of encouraging reporting requires more than scientific confidence: it demands all the advanced skills of persuasion, motivation, and marketing. Some countries have shown great imagination and creativity in this area, but the impact of the science on public health will depend heavily on a much more committed and professional approach to making information relevant and meaningful and to influencing behavior.

Medical therapy becomes ever more complex. Multiple drug use may result in adverse interactions. Not only is there polypharmacy caused by a single physician treating multiple disease processes, but with increasing specialization, more than one doctor may be prescribing without others' knowledge. Moreover, there are the drugs which the patient may use from an ever-increasing selection of over-the-counter drugs and herbals, made available in increasingly sophisticated societies. Treating complex diseases also requires consideration of the interaction between concomitant diseases and drugs used for the target illness.

It is clear that there is more and more pressure on doctors and health professionals in general. The increasing technical and professional complexity of their work is apparent, and we must add to that the increasing administrative and bureaucratic load they have, as well as their gross work overload in many countries. Undergraduate medical training does not give sufficient time to adverse reactions as a most important cause for morbidity; postgraduate education is too frequently concerned with the latest therapy and the importance of being up-to-date in the scholarly rather than the practical sense. There is unending pressure on doctors, including the threat of litigation for even the most genuine of errors by the most careful of doctors. Patients are increasingly informed on medical matters and encouraged, rightly, to understand their therapy and to be active partners rather than passive objects in its management. Unfortunately, the reliability of information sources is very variable, including that massive amount on the Internet. This involves doctors

in an increasing need to justify their advice on therapy and even to undo confusion because of conflicting information.

Good communication practices and the best use of information technology should be very high on the agenda of everyone committed to drug safety improvement. This offers the only way of ensuring that health professionals can easily express their concerns over the safety of drug products to an agency that can collate, analyze, and use the information for focused communication back to health professionals and their patients in ways that can be useful in daily practice, and not be seen merely to add to information overload.

The problem is not just that there is more information to assimilate, but that it is currently provided as isolated, unfocused messages. What we need are focused, relevant messages available at the critical points of need: when doctors are prescribing, pharmacists are dispensing, and patients are being treated.

ACKNOWLEDGMENTS

We gratefully acknowledge the work of Dr Mary Couper in the review of this chapter and her many helpful comments. She has made many of the recent developments within the WHO program possible by her great energy and insights into what needs to be done.

We also acknowledge the work and ideas which come from very many colleagues from the, currently, 73 countries around the world who constantly enliven us with new perspectives and insights.

Key Points

- Global reporting of concerns about suspected adverse drug reactions is a vital alerting tool.
- The reports are a form of “consumer intelligence” containing patient and health professional views on problems relating to treatment. They are not a form of epidemiology.
- Data mining with artificial intelligence (neural networks) allows relationships and patterns to be seen in the data which would otherwise be missed.
- Interpretation of spontaneous reports always requires careful analysis, thought, and clear communication of results, conclusions, and limitations.

SUGGESTED FURTHER READINGS

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.

Bate A, Orre R, Lindquist M, Edwards IR. Explanation of data mining methods. *BMJ* website 2001: <http://www.bmj.com/cgi/content/full/322/7296/1207/DC1>.

Bate AJ, Lindquist M, Edwards IR, Orre R. Understanding quantitative signal detection methods in spontaneously reported data. *Pharmacoepidemiol Drug Saf* 2002; **11** (Suppl 1): 214–15.

CIOMS. *Benefit–Risk Analysis for Marketed Drugs: Evaluating Safety Signals*, 1st edn. Geneva: World Health Organization, 1998.

CIOMS. *Current Challenges in Pharmacovigilance: Pragmatic Approaches*. Geneva: CIOMS, 2001.

DuMouchel W. Bayesian Data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat* 1999; **53**: 177–90.

Edwards IR, Lindquist M, Wiholm BE, Napke E. Quality criteria for early signals of possible adverse drug reactions. *Lancet* 1990; **336**: 156–8.

Edwards IR. Adverse drug reactions: finding the needle in the haystack. *BMJ* 1997; **315**: 500.

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.

Finney DJ. Systematic signalling of adverse reactions to drugs. *Methods Inf Med* 1974; **13**: 1–10.

Lindquist M, Stahl M, Bate A, Edwards IR, Meyboom RH. A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. *Drug Saf* 2000; **23**: 533–42.

Meyboom RH, Egberts AC, Edwards IR, Hekster YA, de Koning FH, Gribnau FW. Principles of signal detection in pharmacovigilance. *Drug Saf* 1997; **16**: 355–65.

Meyboom RHB. Good practice in the postmarketing surveillance of medicines. *Pharm World Sci* 1997; **4**: 19.

Meyboom RHB. The case for good pharmacovigilance practice. *Pharmacoepidemiol Drug Saf* 2000; **9**: 335–6.

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.

UMC. *Safety Monitoring of Medical Products*. Uppsala: Uppsala Monitoring Centre, 2000.

UMC. *Viewpoint*. Uppsala: Uppsala Monitoring Centre, 2002.

van Puijenbroek EP, van Grootheest K, Diemont WL, Leufkens HG, Egberts AC. Determinants of signal selection in a spontaneous reporting system for adverse drug reactions. *Br J Clin Pharmacol* 2001; **52**: 579–86.

van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf* 2002; **11**: 3–10.

Venning GR. Identification of adverse reactions to new drugs. II. How were 18 important adverse reactions discovered and with what delay. *BMJ* 1983; **286**: 289–92.

WHO. *The Importance of Pharmacovigilance. Safety Monitoring of Medicinal Products*. Geneva: World Health Organization, 2002.

9

Case–Control Surveillance

The following individuals contributed to editing sections of this chapter:

LYNN ROSENBERG, PATRICIA F. COOGAN, and JULIE R. PALMER

Slone Epidemiology Center, Boston University, Boston, Massachusetts, USA.

INTRODUCTION

There is no assurance that medications are safe at the time they are released to the market, because premarketing trials for safety and efficacy are too small to detect any but common adverse effects and too brief to detect effects that occur after long latent intervals or durations of use. Indeed, as described in Chapter 1, numerous drugs have been removed from the market, sometimes many years after approval. Postmarketing surveillance serves not only to document unintended adverse effects of medications, but also to document beneficial effects unrelated to the indications for use. Documentation of long-term safety is also important, particularly for drugs that are widely used by healthy individuals. Since drugs used for chronic conditions tend to be taken regularly and for long periods, there may well be unintended health effects. On the other hand, removal of a drug from the market because of concerns that turn out to be unfounded would not serve the public's health.

The need for surveillance of prescription medications is clear. However, non-prescription drugs can also have serious adverse effects and unintended benefits. More and more drugs previously available only by prescription, such as ibuprofen, naproxen, and cimetidine, are being approved for over-the-counter sales, and the change from prescription

to non-prescription sales often results in large increases in use. Until recently, most over-the-counter medications were used for acute self-limiting conditions. However, therapeutic areas that are currently under consideration by pharmaceutical companies for changes from prescription to non-prescription sales include hypercholesterolemia, osteoporosis, hypertension, and depression.

The use of dietary supplements, including herbal supplements, has increased dramatically in recent years. In the Slone Survey, an ongoing survey of a random sample of the US population conducted by the Slone Epidemiology Center, each of 10 supplements had been taken in the preceding week by at least 1% of the population during the years 1998 to 2001. Dietary supplements are often self-prescribed for many of the same reasons that "traditional" prescription and non-prescription drugs are used. Supplements are sold over-the-counter and they do not have to be shown to be efficacious or safe before being marketed. In view of their widespread use, their potential to act as carcinogens, and their possible influence on estrogen action and metabolism, dietary supplements should be monitored for unanticipated effects on the occurrence of cancer and other illnesses.

Cohort studies, such as linkage studies of pharmacy data with outcome data, figure prominently among the

postmarketing strategies currently in use. These studies are useful for monitoring prescription drugs but generally lack information on non-prescription medications and dietary supplements. They are also problematic for the documentation of carcinogenic effects that may occur long after the initiation of drug use. We have developed a surveillance system, Case-Control Surveillance (CCS), which uses case-control methodology to systematically evaluate and detect effects of medications and other exposures on the risk of serious illnesses, principally cancers. CCS includes monitoring of non-prescription drugs and dietary supplements as well as prescription drugs. CCS also includes a biologic component that allows for the assessment of whether genetic polymorphisms modify the effect of a medication or supplement on the risk of the illness.

DESCRIPTION

OVERVIEW

CCS began in 1976 when the US Food and Drug Administration (FDA) provided funding for the monitoring of nonmalignant and malignant illnesses in relation to medication use. Because of concerns about the effects of medications on cancer risk (e.g., postmenopausal female hormone use on risk of endometrial cancer), we sought funding to continue CCS, with a focus on cancers. The National Cancer Institute has provided funding for that purpose since 1988.

In CCS, multiple case-control studies are conducted simultaneously. Individuals with recently diagnosed cancer or nonmalignant conditions are interviewed in a set of participating hospitals. Information is obtained by standard interview on lifetime history of regular medication use and factors that might confound or modify drug-disease associations. Inpatient drug use is generally not recorded. The discharge summary is obtained for all patients, and the pathology report for patients with cancer. A biologic component was added in 1998: participants provide cheek cell samples from which DNA is extracted and stored. Since the beginning of the study, over 70 000 patients have been interviewed, of whom about 25 000 had recently diagnosed primary cancers of various sites.

The CCS database is used for hypothesis testing and discovery. In-depth analyses of the data are carried out to investigate hypotheses that arise from a variety of sources. The data are also "screened" periodically by means of multiple comparisons to discover new associations. Institutional review board approval has been obtained from

all collaborating institutions, and the study complies with all Health Insurance Portability and Accountability Act (HIPAA) requirements. All participants provide written informed consent separately for the interview and for the buccal cell sample.

METHODS

Case and Control Identification and Accrual

The collaborating institutions, located in several geographic areas, have changed over time. The current network, supervised by Dr Brian Strom, consists of seven teaching and community hospitals in Philadelphia. Hospitals in Baltimore, Boston, New York, and other areas have participated in the past.

Specially trained nurse-interviewers employed by CCS interview adult patients aged 21–79 years in collaborating hospitals. The interviewers enroll patients with recently diagnosed cancers or recently diagnosed nonmalignant disorders; the latter serve as a pool of potential controls in case-control analyses, and from time to time a control diagnosis may itself be of interest as the outcome (e.g., cholecystitis, pelvic inflammatory disease). Patients with conditions of acute onset (e.g., traumatic injury, appendicitis) are suitable controls in many analyses, and they are selectively accrued. For more chronic conditions (e.g., orthopedic disorders, kidney stones), recruitment is confined to patients whose diagnosis was made within the previous year. Only patients living in areas within about 50 miles of the hospital are eligible; the interviewers have a list of acceptable ZIP codes and only patients residing in those areas are interviewed. To accrue cases of special interest, the interviewers selectively seek out patients with particular diagnoses according to a priority list. The interviewers find cases through a variety of methods, including checking admissions lists and patient charts. If the interviewer has a choice of interviewing a patient with a priority cancer and a patient with a cancer of another site, the priority cancer will be chosen. The interviewers try to interview all new cases but hospital stays are short and patients are often occupied with having tests, treatments, and visitors. Thus, in practice, the interviewers enroll all patients who are available. Patients are recruited without knowledge of their exposure status. Written informed consent is obtained before interviews are conducted. The interview setting—a hospital or clinic room—is similar for cases and controls. The interviewers are unaware if the patient is a "case" or "control" because many diseases and hypotheses are assessed, and cases in one analysis may be controls in another.

Participation rates in CCS exceeded 90% before the inclusion of the collection of cheek cell samples. After the addition of this biologic component, about 20% of patients have refused to participate. Currently, among the 80% of patients who agree to be interviewed, about 95% provide a cheek cell sample. Patients who agree to provide a biologic sample in addition to the interview are similar in age and sex to those who participate only in the interview; white patients are slightly more likely to participate in the biologic component than black patients.

Table 9.1 shows the numbers of patients with newly diagnosed cancer of various sites that have been accrued in CCS since 1976 in the four largest centers in which CCS has operated—Baltimore, Boston, New York City, and Philadelphia. All subsequent tables refer to the same four areas. CCS currently includes 7160 patients with breast cancer, about 2700 with large bowel cancer, at least 1000 each with lung cancer, malignant melanoma, prostate cancer, or ovarian cancer, and at least 500 each with endometrial cancer, leukemia, bladder cancer, pancreatic cancer, non-Hodgkin's lymphoma, or renal cancer.

Table 9.2 lists the more common diagnoses among patients admitted for nonmalignant conditions. These patients serve as a pool of controls for analyses of various cancers, although in some instances nonmalignant diagnoses themselves have been assessed as the outcome of interest.

Table 9.1. Cases of incident cancer of selected sites accrued in CCS since 1976; Baltimore, Boston, New York City, and Philadelphia

Cancer	No.
Breast	7160
Large bowel	2700
Lung	1770
Malignant melanoma	1495
Prostate	1375
Ovary	1000
Endometrium	870
Leukemia	815
Bladder	600
Pancreas	575
Non-Hodgkin's lymphoma	530
Kidney/kidney pelvis	525
Testis	420
Hodgkin's disease	310
Stomach	345
Esophagus	240
Gallbladder	135
Choriocarcinoma	50
Liver	60
Small intestine	40

Table 9.2. Patients with nonmalignant conditions accrued in CCS since 1976; Baltimore, Boston, New York City, and Philadelphia

Nonmalignant condition	No.
Fracture	2750
Other injury	2700
Uterine fibroid	1700
Benign neoplasm	1520
Cholecystitis	1460
Displacement of intervertebral disc	1360
Ovarian cyst	1350
Hernia	1080
Appendicitis	930
Cholelithiasis	890
Calculus of kidney and ureter	720
Pelvic inflammatory disease	700
Benign prostatic hypertrophy	610
Ectopic pregnancy	470
Diverticulitis	420
Endometriosis	390
Cellulitis	380
Pancreatitis	290
Spinal stenosis	250
Bowel obstruction	250

Among the most common nonmalignant 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.

Interview Data

Drug Information

It is not feasible to ask specifically about thousands of individual drug entities. Instead, histories of medication use are obtained by asking about use for 43 indication or drug categories, e.g., headache, cholesterol lowering, oral contraception, menopausal symptoms, herbals/dietary supplements. The drug name and the timing, duration, and frequency of use are recorded for each episode of use. The drug dose is recorded when it is part of the brand name, e.g., for oral contraceptives and conjugated estrogens, the brand name sometimes indicates the dosage.

Thousands of different specific medications have been reported. Table 9.3 shows the prevalence of reported use of selected medications and drug classes for which there have been marked changes in the prevalence of use over time. The left-hand column shows the prevalence among CCS patients interviewed during 1976–2003, and the right-hand column the prevalence among patients interviewed

Table 9.3. Use of selected drugs and drug classes in CCS, 1976–2003 and 1998–2003; Baltimore, Boston, New York City, and Philadelphia

Category	1976–2003 (n = 61672) (%)	1998–2003 (n = 4317) (%)
Aspirin-containing drugs	47.5	30.5
Oral contraceptives (women only)	42.0	55.5
Acetaminophen-containing drugs	38.9	55.3
Conjugated estrogens (women aged 50+)	23.7	34.6
Benzodiazepines	20.8	12.3
Thiazide diuretics	14.2	9.2
Ibuprofen	13.7	32.7
Phenylpropranolamine	9.6	2.7
Beta-adrenergic blockers	8.7	11.8
Histamine H ₂ antagonists	7.4	16.4
Phenothiazines	4.2	2.6
Calcium channel blockers	3.9	13.0
Oral anticoagulants	3.9	6.7
Phenolphthalein laxatives	3.8	0.7
Aromatic anticonvulsants	3.7	2.5
Naproxen	3.3	8.1
Phenobarbital	2.6	1.0
Indomethacin	2.5	1.1
Insulin	2.3	4.1
Statins	1.6	13.2
Selective serotonin uptake inhibitors	1.3	9.5

Table 9.4. Use of selected drugs in CCS, 1976–2003; Baltimore, Boston, New York City, and Philadelphia

Drug name	%		
Aspirin	32	Levothyroxine sodium	3
Acetaminophen	29	Warfarin sodium	3
Ascorbic acid	20	Propoxyphene HCl	3
Diazepam	14	Norinyl	3
Ibuprofen	14	Oxycodone/APAP	3
Tocopherol acetate	13	Aluminum hydroxide/magnesium hydroxide	3
Iron	13	Methyldopa	3
Tetracycline	8	Cyanocobalamin	3
Ampicillin	7	Percodan®	3
Erythromycin	6	Acetaminophen with codeine	3
Aluminum hydroxide gel	6	Midol®	3
Hydrochlorothiazide	6	Diphenhydramine HCl	2
Cortisone	6	Indomethacin	2
Prednisone	6	Psyllium hydrophilic colloid	2
Guaifenesin	5	Codeine	2
Furosemide	5	Pseudoephedrine HCl	2
Vitamin B complex	4	Potassium chloride	2
Synalgos	4	Chlorpheniramine maleate	2
Propranolol HCl	4	Nitroglycerin	2
Bufferin®	4	Contoz®	2
Cimetidine	4	Sulfisoxazole	2
Calcium	4	Digoxin	2
Triamterene/hydrochlorothiazide	4	Yellow phenolphthalein	2
Miconazole nitrate	4	Nyquil®	2
Chlordiazepoxide HCl	3	Diphenoxylate HCl/atropine SO ₄	2
Vitamin A	3	Milk of magnesia	2
Ranitidine HCl	3	Thyroid	2
		Fiorinal®	2
		Trimethoprim/sulfamethoxazole	2
		Excedrin®	2

during 1998–2003. There were large increases in recent years in the use of acetaminophen, various nonsteroidal anti-inflammatory drugs, histamine H₂ antagonists, calcium channel blockers, statins, and selective serotonin reuptake inhibitors. Some of these increases were attributable in part to changes from prescription to over-the-counter sales; for example, those for the histamine H₂ antagonist cimetidine, and the nonsteroidal anti-inflammatory drug ibuprofen.

In recent years, CCS patients have increasingly reported the use of dietary supplements. In 2000–2003, use of glucosamine was reported by 1.5%, ginkgo biloba by 1.3%, echinacea by 1.2%, and ginseng by 0.9%.

Tables 9.4 and 9.5 show the frequency of use of the most commonly reported drugs and drug classes by CCS participants from 1976 to 2003.

Information on Factors Other Than Drugs

Information on many factors that may confound or modify drug–disease associations is routinely collected: 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),

Table 9.5. Use of selected drug classes in CCS, 1976–2003; Baltimore, Boston, New York City, and Philadelphia

Drug class	%
Vitamins/minerals	62
Aspirin-containing drugs	47
Acetaminaphen-containing drugs	39
Iron	36
Oral contraceptives	26
Folic acid	26
Antihistamines	24
Estrogens	22
Benzodiazepines	21
Corticosteroids	17
Narcotic pain formulas	16
Vitamin A	15
Antacids	15
Thiazides	14
Diazepam	14
Ibuprofen	14
Sulfonamides	13
Laxatives	11
Folic acid antagonists	10
Tetracyclines	10

Phenylpropanolamine	10
Calcium salts	10
Beta-adrenergic blockers	9
Ampicillin/amoxicillin	8
Phenacetin	8
Pseudoephedrine	8
Histamine H ₂ antagonists	7
Macrolide antibiotics	7
Codeine	6
Conjugated estrogens	6
Antifungals	6
Barbiturates	6
Thyroid supplements	6
Guaifenesin	6
Antidepressants	5
Furosemide	5
Phenothiazines	4
Docusate salts	4
Calcium channel blockers	4
Oral anticoagulants	4
Hypnotics and tranquilizers	4
Cephalosporins	4
Aromatic anticonvulsants	4
Tricyclic antidepressants	4
Naproxen	3
Methyldopa	3
Antimalarials	3
Sulfonylureas	3
Nitrates	3
ACE inhibitors	3
Xanthines (excludes caffeine)	3
Phenobarbital	3
Cardiac glycosides	2
Indomethacin	2
Digitalis	2
Insulin	2
Meprobamate	2
Heparin	2
Statins	2
Other anti-hyperlipidemics	2
Aminoglycosides	2

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 two years). These factors may be of interest in their own right as risk factors.

Information from the Hospital Record

A copy of the discharge summary is obtained for every patient enrolled in the study and the pathology report for all patients with cancer. These are reviewed and abstracted in the central office by the study nurse-coordinator, blind

to exposure category, in order to properly classify the diagnosis.

Buccal Cell Samples

The collection of buccal cell samples from patients in CCS began in 1998. Patients who agree to provide samples rub the inside of each cheek with a brush (two samples per patient). This method of DNA collection is suitable for hospital patients because it is noninvasive. The samples are mailed to the collaborating laboratory for extraction and storage of the DNA. Samples collected in this manner have been analyzed successfully for the NAT2-341 gene polymorphism, attesting to the quality and quantity of the extracted DNA.

The stored DNA serves as a resource to identify subgroups that may be at increased risk of particular outcomes related to particular exposures by virtue of inherited genotype, and to elucidate mechanisms of carcinogenesis. The metabolism of environmental carcinogens, including drugs, likely involves genes that regulate phase I monooxygenation and phase II conjugation of potential carcinogens.

Drug Dictionary

Our research group has for many years maintained a drug dictionary. The dictionary is a computerized linkage system composed of individual medicinal agents and multicomponent products, each assigned a specific code number. All combination products are linked to their individual components. Thus, groupings (coalitions) of drugs that contain a particular entity can be easily formed. For example, "Tylenol" is contained in some 50 products coded in our drug dictionary. The constituents of the products can be obtained from the dictionary; e.g., Tylenol Cold Effervescent Formula contains acetaminophen, chlorpheniramine maleate, and phenylpropanolamine HCl. Tylenol products and all other products containing acetaminophen are contained in the acetaminophen coalition, a total of over 450 products. Coalitions of many other types of drugs have also been formed, e.g., selective serotonin reuptake inhibitors, calcium channel blockers, tricyclic antidepressants, thiazide diuretics, benzodiazepines, and beta-adrenergic blockers. The dictionary is continuously maintained and updated by research pharmacists who determine the components of newly encountered products, assign code numbers, and update coalitions. The dictionary currently contains over 15 000 single agent and 7900 multicomponent product codes linked to some 23 000 commercial products, including dietary supplements.

DATA ANALYSIS

Hypothesis Testing

Case and Control Specification

For each analysis, the case series is defined, e.g., women with invasive primary breast cancer diagnosed less than a year before admission and documented in the pathology report. Proper control selection is essential for validity. For the particular exposure at issue, appropriate controls should have been admitted for conditions that are not caused, prevented, or treated by that exposure. Our approach is to select three or four appropriate diagnostic categories with sufficient numbers to allow for examination of uniformity of the exposure of interest across the categories. If our judgment about control selection is correct, the prevalence of that exposure will be uniform across the diagnostic categories selected for that analysis.

Aspects of Drug Use

We assess use that began at least a year before admission because use of more recent onset could not have antedated the onset of the cancer. Depending upon the hypothesis, different categories of drug use are of interest. For example, for breast cancer, analyses may focus on drug use at potentially vulnerable times during reproductive life (e.g., soon after menarche, before the birth of the first child, in the recent past). The particular drug or drug regimen may also be relevant, e.g., 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.

The observation of greater effects for more frequent or long duration use provides support for a causal role. Some drugs, particularly non-prescription drugs such as aspirin, other NSAIDs, and acetaminophen, are often used sporadically. Sporadic use in the past cannot be reported accurately. Furthermore, regular use is more likely to play an etiologic role than sporadic use. Thus, our greatest reliance is placed on regular use (e.g., at least 4 times a week for at least 3 months), and particularly on regular use for several years or more.

The timing of use may also be relevant. In our analysis of CCS data on nonsteroidal anti-inflammatory drugs and large bowel cancer, we found that use that had ceased at least a year previously was unrelated to risk, whereas use that continued into the previous year was associated with a reduced odds ratio. The latter relationship had been suggested by the animal data. In addition, there was no excess of cases among past users, suggesting that cessation of use, possibly due to symptoms, did not explain the inverse association with use that continued into the previous

year. “Latent interval” analyses may focus on whether an effect appears long after use. For example, analyses in our assessment of a non-drug exposure vasectomy in relation to the risk of 10 cancer sites considered the interval between vasectomy and the occurrence of the cancer. Also of interest is how long an increased or reduced odds ratio persists after an exposure has occurred. For example, in our assessment of the risk of ovarian cancer in relation to oral contraceptive use, the reduction in users persisted for 15–19 years after cessation of use, extending the previous period, which had been estimated to be about 10–15 years.

The dose of drugs used in the past is difficult to study because of inaccurate recall. For example, women generally use several different brands of oral contraceptives and they have difficulty remembering the brand (with dosage) accurately. Therefore, we do not ask for the dose of the drug used, although the medication name sometimes indicates the dose. For all drugs, the frequency of use and duration provide a useful measure of the intensity of exposure.

Control of Confounding Factors

Odds ratios (and 95% confidence intervals) are estimated from multiple logistic regression analysis. We first identify potential confounding factors, i.e., risk factors for the disease of interest that are related to use of the drug of interest among the controls. Potential confounding factors are controlled in the regression models if their inclusion materially alters the odds ratio, e.g., by 10% or more.

Effect Modification

Certain subgroups may be particularly vulnerable to or particularly protected by an exposure. 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 supplement use was associated with increased risk of breast cancer among thin women, as observed elsewhere. We generally test for interactions specified *a priori* on the basis of results of previous studies or biologic plausibility.

Statistical Power

CCS has excellent statistical power for the detection of associations that are of public health importance. Table 9.6 shows the sample sizes needed for 80% power to detect a range of odds ratios for a range of exposure prevalences.

Table 9.6. Estimated number of cases for detection of various odds ratios, given various drug exposure prevalences in the controls*

Exposure prevalence in controls (%)	Odds ratios			
	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	17 340	4740	1390	710

*Power = 80%, $\alpha = 0.05$ (two-tailed); control-to-case ratio = 4:1.

Drug/Genotype Analyses

Whether an association between a drug exposure and a cancer is modified by inherited genotype is assessed in two ways: by examining the relation of use of the drug to cancer risk within strata of those with and without the genotype of interest, and by the inclusion of an exposure–genotype interaction term in the logistic regression model.

Discovery of Unsuspected Associations

Animal data may lead to the identification of new associations in CCS data. For example, experiments in rodents suggested that nonsteroidal anti-inflammatory drugs might reduce the occurrence of large bowel cancer. An analysis of CCS data revealed an inverse association of large bowel cancer with aspirin use, an association that has since been confirmed in many subsequent studies. (See Case Example 9.1.)

CASE EXAMPLE 9.1: A CASE-CONTROL ANALYSIS OF CCS DATA TO ASSESS A HYPOTHESIS RAISED IN ANIMAL EXPERIMENTS

Background

- In experiments in rodents, NSAIDs reduced the occurrence of chemically induced colon cancer.

Question

- What is the relation of regular NSAID use to the incidence of colorectal cancer in humans?

(Continued)

Approach

- Identify cases of recently diagnosed colon and rectal cancer confirmed by pathology reports in the CCS database (1326 cases).
- Select controls from among patients admitted to hospital for malignant or nonmalignant conditions unrelated to NSAID use (1011 cancer controls and 3880 noncancer controls).
- Use multiple logistic regression to estimate the odds ratio of colorectal cancer in regular users of NSAIDs relative to never users, controlling for potential confounding factors that included age, sex, geographic area, year of interview, race, religion, coffee and alcohol consumption, history of large bowel cancer in a parent or sibling, and number of previous hospitalizations.

Results

- The multivariate odds ratio for recent regular NSAID use (at least 4 days a week for at least 3 months) relative to never use was 0.5 (95% confidence interval 0.4–0.8), based on 48 case users, 233 control users, 457 case never users, and 1733 control never users.
- The inverse association was apparent among women and men and younger and older patients, and for both colon and rectal cancer, and regardless of whether cancer or noncancer controls were used.
- The odds ratio declined as the duration of use increased but the trend was not statistically significant.

Strengths

- Reporting bias was unlikely: the data were collected before the hypothesis was known, and NSAID use was ascertained in the context of eliciting medication use for many indications.
- Important potential confounding factors were controlled.
- The controls were admitted for conditions unrelated to NSAID use.
- Regular NSAID use was assessed; such use, particularly if long-term, is likely to be well remembered.
- Over-the-counter NSAID use, which is most NSAID use, would have been ascertained, if use was regular.
- Participation rates were high.

Limitations

- NSAID use was not asked about specifically and was undoubtedly underreported.

- Information on some potential confounding factors, such as symptoms that preceded the diagnosis of colorectal cancer, was not obtained; if patients with symptoms had given up NSAID use, this would have produced a spurious reduction in the odds ratio.
- The possibility of selection bias in either the case or control series could not be ruled out.

Summary Points

- NSAID use may reduce the incidence of colorectal cancer.
- The results of the present hypothesis-raising study require confirmation in further studies with varying designs, and that take into account confounding factors not assessed in the present study (e.g., symptoms of colorectal cancer, use of screening endoscopy).
- Effects of NSAIDs could vary by type of NSAID, frequency of use, and dose, making it important to assess these aspects of use in further studies.
- A randomized trial of NSAIDs and colorectal cancer would be desirable; however, such a study would need to be very large because of the relatively low incidence of colorectal cancer and would therefore be extremely costly, making such a study unlikely.
- Colorectal polyps are precursors of colorectal cancer; a randomized trial of NSAIDs and polyps would be feasible.

Associations are also identified by systematic “screening” of the data, in which the prevalence of use of a particular drug or drug class (standardized for age, sex, and hospital) among patients with a particular cancer or other illness of interest is compared with the prevalence among patients with other illnesses. Often significant associations ($p < 0.05$) seen in a screen disappear once further cases and controls are enrolled in CCS, or after analyses in which there is careful specification of the case and control groups and control for confounding factors other than age, sex, and study center. We carry out in-depth analyses of new associations if the association is replicated in data collected in CCS in subsequent years, is explained by a highly plausible mechanism, or is of public health importance. Non-drug factors are also screened, and it was in the course of such a screen that we observed an unexpected association between alcohol use and breast cancer. Examples of other unexpected associations from screening are oral contraceptive use with choriocarcinoma and with Crohn’s disease. All of these associations have received independent confirmation.

Further evidence for the validity of the screen findings is the appearance of many known associations that were discovered previously, such as the increased risks of myocardial infarction and venous thromboembolism associated with oral contraceptive use.

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 are presented with the utmost caution.

STRENGTHS

ASSESSMENT OF NON-PRESCRIPTION MEDICATIONS AND DIETARY SUPPLEMENTS AS WELL AS PRESCRIPTION MEDICATIONS

CCS can be used to test hypotheses concerning use of all reported prescription medications from any source. Monitoring systems that rely on pharmacy data can assess only those medications that are prescribed within the system; prescriptions obtained elsewhere (e.g., family planning clinics, friends, and relatives) cannot be assessed. Sometimes prescribed medications are not taken, which is a disadvantage of relying on prescription data.

CCS is the only surveillance system that systematically assesses use of non-prescription products, both non-prescription medications and dietary supplements. The prevalence of dietary supplement use has become high enough that assessment of their effects on disease occurrence is of public health importance.

CCS has documented adverse effects of medications, such as increased risk of liver cancer and breast cancer associated with oral contraceptive use, and increased risk of localized and advanced endometrial cancer associated with postmenopausal estrogen supplement use. Protective effects have also been documented with CCS data, e.g., oral contraceptive use related to reduced risks of ovarian and endometrial cancer, and aspirin use associated with reduced risks of colorectal cancer and stomach cancer. CCS has often documented the safety of drugs after alarms were raised about adverse effects. For example, in experiments in rodents given phenolphthalein, an agent used in non-prescription laxatives, there were increased risks of several cancers. The FDA called for human data on this question. CCS responded and found no increased risk. A small cohort study suggested that calcium channel blockers increased the risk of several cancers; results from the much larger CCS database refuted that finding. Animal data raised the concern that benzodiazepines increased the risk of several cancers;

data from CCS were null. Animal data raised the possibility of increased risks of cancer associated with hydralazine use; CCS results were null.

Many case-control or cohort studies have reported on selected medications, such as noncontraceptive estrogens or oral contraceptives, but comprehensive information on a wide variety of drugs is not routinely collected. The effects of many drugs have not been well assessed. CCS has provided data on the risk of various outcomes in relation to a wide range of medications, including ACE inhibitors, acetaminophen, antidepressants, antihistamines, aspirin and other NSAIDs, benzodiazepines, beta-androgenic blockers, calcium channel blockers, female hormone supplements, hydralazine, oral contraceptives, phenolphthalein-containing laxatives, phenothiazines, rauwolfia alkaloids, selective serotonin reuptake inhibitors, statins, thiazides, and thyroid supplements.

DISCOVERY OF UNSUSPECTED ASSOCIATIONS

Because CCS obtains data on many exposures and many outcomes, the system has the capacity for discovery of unsuspected associations. For example, an inverse association between aspirin use and risk of colorectal cancer was documented in CCS. The publication of the finding provoked many subsequent studies, which confirmed the association. The National Cancer Institute found the findings to be of sufficient potential public health importance to support a randomized trial of aspirin as a preventive of colonic polyps. Other associations discovered in CCS are positive associations of long-term oral contraceptive use with gestational trophoblastic disease and with Crohn's disease, and of alcohol consumption with increased risk of breast cancer. These associations have been confirmed in subsequent studies.

ASSESSMENT OF EFFECTS AFTER LONG INTERVALS OR DURATIONS OF USE

Because the effects of drugs, particularly carcinogenic effects, may become evident only after many years, the capacity for a surveillance system to assess long latent intervals or long durations of use is important. The case-control design used by CCS is efficient for assessing the effects of exposures that occurred in the distant past or after long durations of exposure. For example, CCS documented that the adverse effect of estrogen supplements on risk of endometrial cancer persisted for 15–19 years after cessation of use. Cohort studies are ill suited for these assessments unless the study has been collecting information for many years.

CONTROL OF CONFOUNDING

In observational research, control of confounding is crucial for validity. Drug use is a health-related activity and is associated with factors such as medical history that are in turn strongly associated with disease risk. CCS systematically collects detailed information on important potential confounding factors. These include demographic characteristics, aspects of medical history, reproductive and gynecologic history, family history of cancer, use of tobacco and alcohol, and use of medical care, in addition to use of prescription and non-prescription drugs and dietary supplements. Thus, it is possible to control for these factors in multivariable analyses.

ACCURATE OUTCOME DATA

For all patients, CCS collects information from the hospital record. Pathology reports are obtained for all patients with cancer. CCS is therefore able to accurately classify the diagnosis for which the patient was admitted.

HIGH STATISTICAL POWER

CCS has accrued a large database, with large numbers of patients with cancers of various sites and other illnesses (Table 9.1). Many drugs or drug classes have been taken by at least 1% of the population (Tables 9.4 and 9.5). CCS has high statistical power relative to cohort studies, with excellent power to assess the effects of exposures of public health importance. As shown in Table 9.6, small odds ratios associated with uncommon drug exposures can be detected for common cancers. For less common cancers, odds ratios associated with more common exposures can be detected. For very rare cancers, only relatively large effects can be detected for relatively common exposures. However, an appreciable number of cases of a rare cancer will be attributable to the exposure only when the odds ratio is large and the exposure common.

BIOLOGIC COMPONENT

Unanticipated adverse or beneficial effects of medications may be confined to vulnerable subgroups. CCS has the capacity to assess whether those subgroups are defined by genetic polymorphisms, i.e., whether genetic polymorphisms modify drug-disease associations. This can serve both to identify vulnerable populations and to elucidate mechanisms.

PRODUCTIVITY AND SUBSTANTIVE FINDINGS

CCS has been highly productive: 79 papers have been published. Some of the associations assessed have been briefly described in this chapter.

WEAKNESSES

POTENTIAL FOR BIAS

Selection Bias

When feasible, population-based case-control studies (i.e., identifying all cases in a geographic region and a random selection of non-diseased from the same population as controls) are optimal. Population-based CCS is infeasible for logistic and budgetary reasons. Even in population-based data, however, biased selection of cases and controls may occur because of non-participation. In CCS, the high participation rates reduce the potential for selection bias due to non-participation. In addition, the cases in CCS are persons with various cancers admitted to the hospitals under surveillance, and they define a secondary base which comprises members of the population at large who would be admitted to the same hospitals were they to develop cancer. Enrollment is limited to cases and controls who live within approximately 50 miles of the hospital. The purpose is to include only persons from the secondary base and to exclude referrals from outside that base. Of course, referral patterns for different cancers could be different.

In the analysis of the risk of a particular cancer, we often select a control group of patients with other cancers judged to be unrelated to the exposure; such controls are probably representative of the same base as the cases. A second control group admitted for nonmalignant conditions guards against the possibility that the exposure may cause all cancers. We check for uniformity of the exposure of interest across the various control categories; uniformity suggests the absence of selection bias.

As another check for bias, a disease unrelated to the drug exposure at issue may be included in the assessment of the relation of that drug to the outcome of interest. For example, in our assessment of acetaminophen in relation to risk of transitional cell cancers, we also assessed renal cell cancer because the latter outcome had not been associated with acetaminophen use.

Recall Bias

It would be desirable to obtain exposure data based on complete and accurate records, with the caveat that people

often do not fill prescriptions or take the drugs prescribed. Validation studies of self-reported prescription drug use are generally difficult in the US because people get drugs from many sources, records are often absent, and participation rates may be suboptimal. Because we believe that recent or long-term use is best remembered, we focus on these categories. The literature on validation of drug use indicates that recent and long-term use of oral contraceptives and female hormone supplements is reported with acceptable accuracy; the product names are less well reported. The relatively few validation studies of other prescription drugs have yielded variable results, with the best agreement for drugs used on a long-term basis, such as those for diabetes and hypertension. A review of validation studies concluded that reporting is affected by the type of medication and drug use patterns (e.g., better reporting for chronically used prescriptions) and by the design of data collection. (See also Chapter 15.) For non-prescription drugs and dietary supplements, validation is infeasible because records of use do not exist.

CCS reduces reporting bias (i.e., differential reporting by cases and controls) by using the same highly structured interview and similar interview settings for cases and controls. Patients are asked about 43 indications for drug use and drug classes. This approach masks hypotheses about particular drugs. Furthermore, control patients admitted for a serious nonmalignant condition are as likely to carefully search their memories as case patients admitted for a cancer. As a check for reporting bias, we may assess a drug or drug class unrelated to the outcome. For example, beta-blockers and ACE inhibitors were assessed in our analysis of cancer risk in relation to calcium channel blockers, because the former drug classes had not been linked to cancer risk.

Nondifferential Misclassification

Nondifferential underascertainment of drug use will weaken observed associations. As an example, the effect on “true” odds ratios of 3, 2, and 1.5 of 30% underascertainment of drug use among cases and controls for a range of “true” exposure prevalences in the controls is given in Table 9.7. While the effect of nondifferential underascertainment is for the estimate to move towards the null, the changes are small, no more than about 10% in the worst case. Effects are of course smaller if underascertainment of drug use is less than 30%. Thus, nondifferential misclassification likely has only small effects on odds ratios estimated for exposures of interest in CCS.

The ultimate test of validity of CCS results is whether they are confirmed by well-conducted studies that use different methods. CCS results have repeatedly passed that test.

Table 9.7. Observed odds ratio given 30% underascertainment of drug use in cases and controls

True prevalence of drug use in controls			
		<i>True odds ratio</i>	
		3.0	1.5
		<i>Observed odds ratio</i>	
15%	2.7	1.9	1.5
10%	2.8	1.9	1.5
5%	2.9	2.0	1.5
1%	3.0	2.0	1.5

PARTICULAR APPLICATIONS

CCS has the capacity to assess the risk of illnesses in relation to use of prescription drugs, non-prescription drugs, and dietary supplements reported by participants. As described in previous sections, CCS has documented increased risk, decreased risk, and absence of risk. In addition, CCS has generated important new hypotheses, probably the most important of which are the positive association of alcohol with breast cancer and the inverse association of non-steroidal anti-inflammatory drugs with large bowel cancer. Now that dietary supplement use has become widespread, CCS will assess the unintended health effects of these agents. When particular issues arise, the system can be steered to selectively accrue cases of the disease of interest, but extremely rare diseases are beyond the scope of CCS and other routine monitoring systems.

The scope of CCS is broad, with major contributions having been made to the evaluation of the health effects of a wide range of medications in relation to a range of illnesses. In recent years, there has been a particular focus on non-prescription medications, such as the widely used nonsteroidal anti-inflammatory drugs, mostly obtained over-the-counter. The diseases assessed include breast cancer, ovarian cancer, endometrial cancer, choriocarcinoma, prostate cancer, large bowel cancer and other gastrointestinal cancers, lung cancer, melanoma, liver cancer, pelvic inflammatory disease, cholecystitis, and venous thromboembolism. The drugs and drug classes assessed include ACE inhibitors, acetaminophen, antidepressants, antihistamines, aspirin and other NSAIDs, benzodiazepines, beta-blockers, calcium channel blockers, female hormone supplements, hydralazine, oral contraceptives, phenolphthalein, phenothiazines, rauwolfia alkaloids,

statins, thiazides, and thyroid supplements. CCS has also made contributions to assessment of the health effects of non-drug factors, such as the tar and nicotine content of cigarettes, menthol cigarette smoking, alcohol and coffee consumption, and vasectomy.

THE FUTURE

Medication use in the US is widespread and increasing, spurred in part by direct marketing to consumers. New prescription drugs continue to be introduced to the market. Until medications and supplements have been used by appreciable numbers of people for appreciable periods, their health effects cannot and will not have been adequately monitored.

CCS will continue to monitor the effects of prescription drugs. The switch from prescription to over-the-counter sales has increased in recent years, and the use of dietary supplements has become widespread. CCS will carry out the monitoring of new and older over-the-counter medications, and of dietary supplements.

Several medications are of particular interest. Statins, the first of which was introduced to the market in 1987, are among the most widely used drugs in the US. Data from *in vitro* experiments suggest that the statins may have chemopreventive potential at various sites, but there is also concern about a potential to increase cancer risk. Selective serotonin reuptake inhibitors are also widely used, often by healthy persons. A recent report of three cases of breast neoplasia among men who took SSRIs raises the concern that these drugs may affect breast cancer incidence. Histamine H₂ antagonists may have a stimulatory effect on the immune system. It has been suggested that cimetidine could prevent prostate cancer, but there are also concerns about possible increases in risk of breast cancer. Nonsteroidal antiinflammatory drugs also require continuing attention because of their widespread use, and because of the introduction of new agents. The inverse association of use with risk of colon cancer has raised interest in assessment of potential effects at other cancer sites. The health effects of dietary supplements are almost entirely unknown; CCS will devote considerable attention to their relation to the risk of cancer.

Knowledge about the actions of genetic polymorphisms has increased greatly in recent years. Genes with allelic variability that regulate the metabolism of drugs are likely candidates for modification of drug–cancer relationships. CCS will have the capacity to assess plausible hypotheses that arise in the future about modification of drug effects on cancer risk by genetic polymorphisms.

ACKNOWLEDGMENTS

CCS was originated in 1975 by Dr Samuel Shapiro and the late Dr Dennis Slone. It was originally supported by contracts from the US Food and Drug Administration. Since 1988, CCS has been supported by the National Cancer Institute (CA45762). Additional support for data analyses has been provided by various pharmaceutical companies, which are acknowledged in the papers that relied on their support.

Key Points

- In Case–Control Surveillance (CCS), multiple case–control studies are conducted simultaneously in order to monitor the effects of prescription and over-the-counter medications and dietary supplements (e.g., herbals) on risk of various illnesses.
- CCS relies on self-reports of medication and dietary supplement use.
- Because it is infeasible to ask about thousands of medications and dietary supplements specifically, CCS asks about use for 43 indications or medication categories.
- CCS has the capacity to monitor not just prescription drugs, but also many potentially important confounders, over-the-counter drugs, and dietary supplements, which can only be ascertained through self-report.
- Because chronic and regular use of medications is better reported than sporadic or short-term use, the main focus of CCS is on regular and long-term use.
- CCS can have high statistical power because of the large number of cases of various illnesses accrued.
- A limitation of hospital-based case–control studies like CCS is the potential for selection bias.

SUGGESTED FURTHER READINGS

- Bayer Corporation. Market withdrawal of Baycol. August 8, 2001. Available at: www.baycol-rhabdomyolysis.com/baycol_recall.
- Brewer T, Colditz G. Postmarketing surveillance and adverse drug reactions: current perspectives and future needs. *JAMA* 1999; **281**: 824–9.
- Coogan PF, Rosenberg L, Palmer JR, Strom BL, Zauberg 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.
- FDA. List of drug products that have been withdrawn or removed from the market for reasons of safety or effectiveness. 21 CFR Part 216. Docket No. 98N-0655. *Fed Reg* 1998; **63**: 54082–9.
- FDA issues voluntary removal of drugs with phenylpropanolamine. Release #00B-139, November, 2000.

- Francesco International. Potential government high priority therapeutic areas for OTC. SWITCH® Newsletter, SwitchTrends. Available at: <http://www.rxtooteswitch.com>. Accessed January 14, 2004.
- 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.
- 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.
- Harlow SD, Linet MS. Agreement between questionnaire data and medical recall: the evidence of accuracy of recall. *Am J Epidemiol* 1989; **129**: 233-48.
- 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.
- Kessler DA. Cancer and herbs. *N Engl J Med* 2000; **342**: 1762-3.
- Lasser KE, Allen PD, Woolhandler SJ, Himmelstein DV, Wolfe SM, Bor DH. Timing of new black box warnings and withdrawals for prescription medications. *JAMA* 2002; **287**: 2215-20.
- Newall CA, Anderson LA, Philpson JD. *Herbal Medicine: A Guide for Healthcare Professionals*. London: The Pharmaceutical Press, 1966.
- Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 1982; **116**: 114-22.
- 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.
- Rosenberg L, Louik C, Shapiro S. Nonsteroidal antiinflammatory drug use and reduced risk of large bowel carcinoma. *Cancer* 1998; **82**: 2326-33.
- Soller RW. OTCS 2000: Achievements and challenges. *Drug Inf J* 2000; **34**: 693-701.
- Thun MJ, Namboodiri MM, Heath CW Jr. Aspirin use and reduced risk of fatal colon cancer. *N Engl J Med* 1991; **325**: 1593-6.
- Thun MS, Namboodiri MM, Calle EE, Flanders MD, Heath CW Jr. Aspirin use and risk of fatal cancer. *Cancer Res* 1993; **53**: 1322-7.

Prescription-Event Monitoring

Edited by:

SAAD A.W. SHAKIR

Drug Safety Research Unit, Southampton, UK.

INTRODUCTION

The thalidomide disaster, which caused the development of 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. The reasons for monitoring postmarketing drug safety were summarized 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):

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.

Premarketing clinical trials are effective in studying the efficacy of medicines. However, while they define many aspects of the safety profiles of medicines, premarketing clinical trials have limitations in defining the clinically necessary safety profiles of drugs. These limitations include:

- the small numbers of patients, in epidemiologic terms, included in premarketing clinical trial programs;
- the large numbers of patients in these programs who receive the study products for short durations (many only receive a single dose); this limits the power of premarketing clinical trials to detect rare ADRs or ADRs with long latency;
- premarketing development programs are dynamic; doses and formulations can change during drug development; in some programs, large numbers of patients studied receive lower doses or different formulations from those eventually marketed;
- the exclusion from clinical trials of special populations such as the young, the old, women of childbearing age, and patients with concurrent diseases eliminates many patients who may be at higher risk for developing ADRs, limiting the generalizability of the results of such trials.

Therefore, there has been general agreement for more than 30 years that the clinically necessary understanding of drug safety depends on postmarketing monitoring and postmarketing safety studies. This has resulted in not only the establishment of voluntary systems for reporting suspected ADRs (see Chapters 7 and 8) but the development of a

range of other methods to monitor and study postmarketing drug safety.

Soon after the establishment of spontaneous reporting systems, it was recognized that, while such systems have many real advantages for detecting and defining ADRs, particularly rare ADRs, they also have limitations.

The theoretical basis for establishing a system to monitor events regardless of relatedness to drug exposure was proposed by Finney in 1965. This and the limited contribution of the spontaneous reporting system in detecting hazards such as the oculomucocutaneous syndrome with practolol led Inman to establish the system of Prescription-Event Monitoring (PEM) at the Drug Safety Research Unit (DSRU) at Southampton in 1981. Subsequently the CSM, wishing to consider monitoring the postmarketing safety of medicines, established a committee under the chairmanship of Professor David Grahame-Smith. The committee reported in June 1983 and again in July 1985, and in these reports showed an appreciation of the need for prescription-based monitoring. It also specifically recommended that postmarketing surveillance (PMS) studies should be undertaken "on newly-marketed drugs intended for widespread long-term use." PEM is one form of pharmacovigilance that, with the development and harmonization of drug regulation in the European Community, has its basis in Directives 65/65 and 75/319, and in Regulation 2309/93.

DESCRIPTION

The PEM process is summarized in Figure 10.1. In the UK, virtually all patients are registered with a National Health Service (NHS) general practitioner (GP), who provides primary medical care and acts as a gateway to specialist and hospital care. The file notes in general practice in the UK include not only information obtained in primary care but data about all contacts with secondary and tertiary care, such as letters from specialist clinics, hospital discharge summaries, and results of laboratory and other investigations. It is a lifelong record; when a patient moves to a new area, all his notes are sent to his new GP.

The GP issues prescriptions for the medicines he/she considers medically warranted. The patient takes the prescription to a pharmacist, who dispenses the medication and then sends the prescription to a central Prescription Pricing Division (PPD) which is part of the NHS-Business Service Authority (NHS-BSA) for reimbursement. The DSRU is, under longstanding and confidential arrangements, provided with electronic copies of all those prescriptions issued throughout England for the drugs

being monitored. Products that are selected for study by PEM are new drugs which are expected to be widely used by GPs; in some cases the DSRU is unable to study suitable products because of limited resources. In addition, the DSRU conducts studies on established products when there is a reason to do so, for example, a new indication or extending usage to a new population. Collection of the exposure data usually begins immediately after the new drug has been launched. These arrangements operate for the length of time necessary for the DSRU to collect the first 50 000 prescriptions that identify 20 000–30 000 patients given the new drug being monitored. For each patient in each PEM study, the DSRU prepares a computerized longitudinal record in date order of the use of the drug. Thus, in PEM, the exposure data are national in scope and provide information on the first cohort to receive the drug being monitored after it has been launched into everyday clinical usage. The exposure data are of drugs both prescribed and dispensed, but there is no measure of compliance.

After an interval of 3–12, but usually 6, months from the date of the first prescription for each patient in the cohort, the DSRU sends to the prescriber a "green form" questionnaire seeking information on any "events" that may have occurred while the patient was taking the drug or in the months that followed. This takes place on an individual patient basis. To limit the workload of GPs, no doctor is sent more than four green forms in any one month. The green form, illustrated in Figure 10.2, is intended to be simple. It requests information on age, indication for treatment, dose, starting date and stopping date (duration of treatment), reasons for stopping therapy, all events which have occurred since the start of treatment, and the cause(s) of death if applicable. The green form includes the definition of an "event," which 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."

A recent development in the PEM process is the inclusion of a small number of "additional" questions in the green forms. Such questions aim to examine aspects such as confounding by indication, concurrent illnesses, and concomitant medications. For example, the green forms in the PEM studies of the COX-2 inhibitors, e.g., celecoxib, included questions regarding previous history of dyspeptic conditions, and the green forms for the PEM studies of PDE5 inhibitors for erectile dysfunction, such as sildenafil, included questions about history of cardiovascular disease.

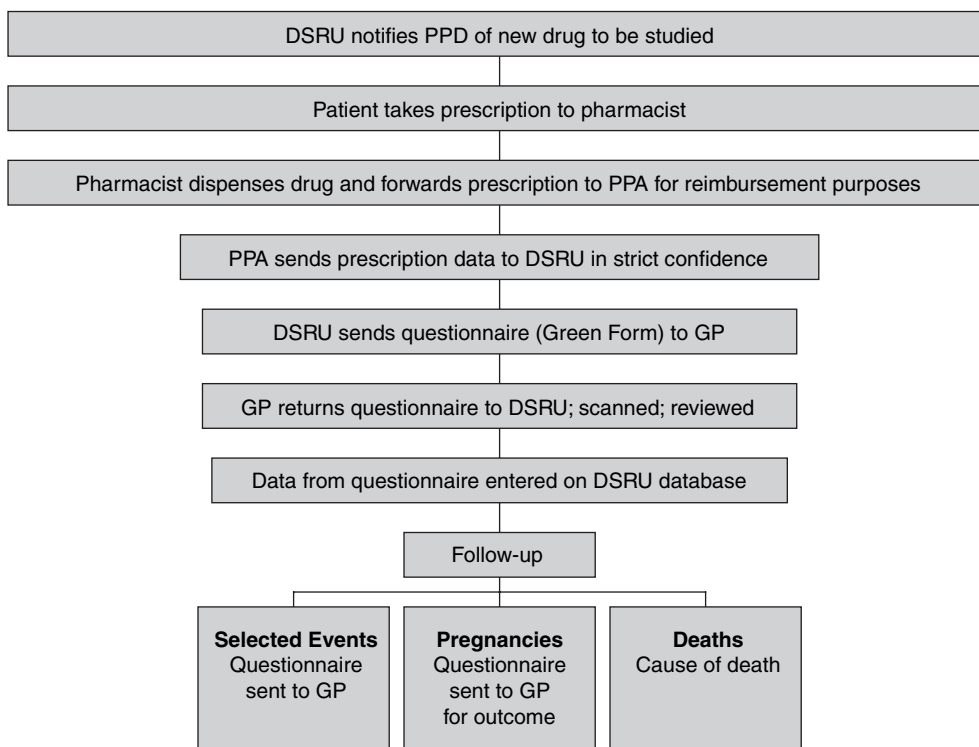


Figure 10.1. The PEM process.

The GP is not paid to provide information to the DSRU, which is provided, under conditions of medical confidence, in the interest of drug safety. The system provides good contact with the GPs and facilitates the collection of any follow-up or additional data considered necessary by the research scientists/physicians monitoring each study and working within the DSRU. Table 10.1 includes a list of the categories of medical events for which follow-up is sought by research fellows. Table 10.2 lists the medically serious events that have been associated with the use of medicines; follow-up information is sought for these too. All pregnancies reported during treatment or within three months of stopping the drug are followed up to determine the outcome.

PEM collects event data and does not ask the doctor to determine whether any particular event represents an ADR. If, however, an event is considered to be an ADR or has been reported by means of the “yellow card” scheme, then the doctor is asked to indicate this on the green form.

Each green form is seen 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, e.g., external communications or expedited follow-up.

Events are coded and entered into a database using a hierarchical dictionary, arranged by system–organ class with specific “lower” terms grouped together under broader “higher” terms. The DSRU dictionary has been developed over the past 20 years and contains 11 640 doctor summary terms (as near as possible to the term used by the reporting doctor, e.g., crescendo angina) and 1720 lower-level terms mapped to 1185 higher-level terms, within 27 system–organ classes.

Interim analyses of the computerized data are usually undertaken every 2500 patients in each study and contacts are, whenever possible, maintained with the company holding the product license, so that the pharmaceutical companies (although the study is independent of them) can comply with the drug safety reporting procedures of the regulatory authorities.

Based on data from 88 PEM studies conducted to date, the GP response rate (percentage of green forms returned) has been $56.0\% \pm \text{SD } 8.3\%$. The mean cohort size has been

Figure 10.2. Green form for the PEM study on Celebrex (celecoxib).

Table 10.1. Events for which follow-up information is sought from GPs

-
- Medically important adverse events reported during premarketing development
 - Medically important events reported during postmarketing in other countries (for products launched elsewhere before the UK)
 - Medically important events considered to be possibly associated with the product during the PEM
 - All pregnancies
 - Any deaths for which the cause is not known or which may be related to the medication
 - Reports of overdose and suicide
-

Table 10.2. 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
 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 Control Agency (MCA), UK.

10 942 patients. The collection periods (the time for which it has been necessary to collect prescriptions yielding a finished cohort size averaging over 10 000 patients) vary markedly depending on the usage of the drug.

The DSRU is an independent registered medical nonprofit organization associated with the University of Portsmouth. The Unit is extensively supported by donations and grants from the pharmaceutical industry. The drugs to be monitored are chosen by the DSRU, preference being given to innovative medicines intended for widespread use. A list of all completed PEM studies is available on the DSRU's website (www.dsru.org).

STRENGTHS

PEM has a number of important strengths. First, as indicated above, the method is non-interventional in nature and does not interfere with the treatment the doctor considers as most appropriate for the individual patient. Information is collected after the prescribing decision has been made and implemented. This means that in PEM, data are collected on patients who would receive the drug in question in everyday clinical practice and not upon some highly selected group of patients who may be nonrepresentative of the "real world" population. In this way, the system avoids the problem of generalizability inherent in randomized clinical trials, including many postmarketing safety clinical trials.

Second, the method is national in scale and provides "real world" data showing what actually happens in everyday clinical practice. It largely overcomes the problem of making clinical trial data truly representative of the whole population that will receive the drug. For example, PEM studies include unlicensed and unlabelled prescribing, e.g., unlicensed prescribing for children.

Third, as indicated above, PEM exposure data are derived from dispensed prescriptions. Considering the large number of patients who do not get a prescription dispensed, this is an advantage compared to pharmacoepidemiologic databases that rely on prescription data.

Fourth, because the data are concerned with events, the method could detect adverse reactions or syndromes that none of the reporting doctors suspect to be due to the drug. The database allows the study of diseases as well as drugs. Both of these advantages are in line with the early proposals of Finney (1965) on event reporting.

Fifth, the method allows close contact between the research staff in the DSRU and the reporting doctors. This facilitates follow-up of important events, pregnancies, deaths, etc. (Tables 10.1 and 10.2), and allows for the maximum clinical understanding of confounders and biases, and the natural history of ADRs.

Sixth, the method prompts the doctor to fill in the green form and does not rely on the clinician taking the initiative to report. This “prompting” effect of PEM is most important; two studies have demonstrated that ADR reporting is more complete in PEM than in spontaneous ADR reporting systems, such as the yellow card system in the UK.

Seventh, the 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. (See Case Example 10.1.)

CASE EXAMPLE 10.1: PEM AND ORAL CONTRACEPTIVES

Background

- Yasmin is a combined oral contraceptive (COC) containing ethinylloestradiol and a new progestogen, drospirenone; it was launched in the UK in May 2002.
- While the association between oestrogen-containing oral contraceptives and venous thromboembolism (VTE) is well established, VTE is a rare event in young women and the risk associated with a new COC cannot generally be determined from clinical trials.
- The initial prescribing information for Yasmin stated that “It is not yet known how Yasmin influences the risk of VTE compared with other oral contraceptives.”

Question

- How can Prescription-Event Monitoring (PEM) 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.

Strength

- 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.

Summary Points

- While premarketing clinical trials identified many aspects of the safety of Yasmin, apparently no cases of VTE were reported.
- An association between the COC and VTE has been recognized for more than 40 years, so it is important to know whether a new COC is associated with VTE and to what extent.
- The lack of selection bias and large numbers of women studied led to identification of women who developed VTE while taking Yasmin, all of whom had risk factors 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 be examined by other studies.

Eighth, the method identifies patients with adverse drug reactions who can be studied further, for example, in nested case-control studies to examine risk factors for ADRs including pharmacogenetic risk factors (see Chapter 18). Relatedly, while information on some co-prescribed drugs

can be obtained in the initial green form, more detailed information about concomitant medications can be obtained for selected cases, e.g., important medical events, during follow-up.

Ninth, the large number of completed PEM studies allows comparisons of the safety profiles of drugs in the same therapeutic groups. It is also possible to conduct comparisons with external data.

Finally, pharmacoepidemiologic methods are complementary. PEM can evaluate signals generated in other systems or databases. Similarly, it provides a technique that can generate signals or hypotheses which can themselves be refuted or confirmed by other pharmacoepidemiologic methods.

WEAKNESSES

Like all pharmacoepidemiologic methods, of course PEM has weaknesses. First, not all of the green forms are returned and this could induce a selection bias.

Second, PEM depends on reporting by doctors. As such, it can be as good as but no better than the clinical notes of the GPs and depends on the accuracy and thoroughness of the doctors in completing the green forms. Underreporting, including underreporting of serious and fatal adverse events, is possible in PEM.

Third, PEM is currently restricted to general practice. Drugs which are mainly used in hospitals cannot be studied with the current method of PEM.

Fourth, while studying exposure by dispensing rather than prescribing is an advantage, there is no measure of compliance using dispensed prescriptions, i.e., it is not known whether the patient actually took the dispensed medication.

Finally, detection of rare ADRs is not always possible even with cohorts of 10 000–15 000 patients.

PARTICULAR APPLICATIONS

SEARCHING FOR SIGNALS

Signal detection and evaluation is the primary concern of pharmacovigilance. Several methods are applied for signal detection in PEM.

Assessment of Important Adverse Events

The initial evaluation is conducted by manual examination by research fellows of newly received green forms

for adverse events that may possibly be related to drug exposure. The assessments of individual reports or clusters of reports take into consideration a number of points, including:

- the temporal relationship (time to onset);
- the clinical and pathological characteristics of the event;
- the pharmacological 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 characteristics, including previous medical history, such as history of drug allergies, presence of renal or hepatic impairment, etc.;
- the possibility of drug interactions.

In this activity, PEM is functioning in a manner very similar to spontaneous reporting systems (see Chapters 7, 8, and 17), although with much higher rates of reporting. An example of a safety signal generated in PEM as a result of careful clinical evaluation is the visual field defect with the antiepileptic drug vigabatrin.

Medically Important Events

As mentioned above, special consideration is given to the categories and events listed in Tables 10.1 and 10.2.

Reasons for Stopping the Drug

The green form questionnaire asks the doctor to record the reason why the drug was withdrawn if, in fact, it was withdrawn. Clinical reasons for stopping a drug are ranked according to numbers received in a list, which 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. The clinical reasons for withdrawal are ranked according to the number of reports of each event and are used to generate signals. For example, PEM has identified drowsiness/sedation and weight gain with the antidepressant mirtazapine, and assessed the strengths of signals generated by other methods in PEM.

Table 10.3. All events reported on green forms for meloxicam (summarizing all reports received throughout a typical PEM study, whether or not the patient was still on the drug)

Event	Total	Mth 1	Mth 2	Mth 3	Mth 4	Mth 5	Mth 6	Mths 1–6	Not known
Denominator total ^a	130 615	19 083	19 075	19 068	19 063	19 054	19 046		
Denominator male	41 711	6172	6167	6166	6164	6160	6158		
Denominator female	86 690	12 586	12 583	12 577	12 574	12 568	12 562		
<i>Skin</i>									
Acne	13	1	1	1	3	1	3	10	–
Acne	8	1	–	1	1	–	3	6	–
Acne rosacea	5	–	1	–	2	1	–	4	–
Alopecia	1	–	–	–	1	–	–	1	–
Cyst sebaceous	14	4	3	2	1	1	1	12	–
Dermatitis	43	8	6	2	3	5	4	28	–
Dermatitis contact	4	2	–	–	1	1	–	4	–
Dry skin	21	5	3	2	3	2	1	16	–
Eczema	91	14	12	10	13	7	9	66	–
Eczema	71	10	9	8	9	5	7	48	–
Eczema atopic	1	1	–	–	–	–	–	1	–
Intertrigo	16	2	3	1	4	2	2	14	–
Pompholyx	3	1	–	1	–	–	–	2	–
Eczema varicose	11	2	2	1	1	1	2	9	–
Eruption bullous	8	–	2	1	1	1	1	6	–
Blister	7	–	2	1	1	–	1	5	–
Pemphigoid	1	–	–	–	–	1	–	1	–
Erythema	3	1	1	–	–	–	–	2	–
Erythema multiforme	1	–	1	–	–	–	–	1	–
Folliculitis	6	–	1	1	2	–	2	6	–
Granuloma	1	1	–	–	–	–	–	1	–
Granulomatosis	1	–	–	–	–	1	–	1	–
Haematoma nail	1	–	–	–	–	–	1	1	–
Hair loss	8	1	1	2	–	–	1	5	–
Herpes simplex, skin	4	–	1	–	1	1	–	3	–
Herpes zoster	42	7	4	7	6	8	6	38	–
Hyperkeratosis	6	–	3	1	–	1	–	5	–
Hyperkeratosis	2	–	1	–	–	1	–	2	–
Pityriasis	4	–	2	1	–	–	–	3	–
Infection skin, unspecified/local bacterial	188	29	21	27	30	25	22	154	–
Abscess skin	31	3	–	4	12	6	3	28	–
Cellulitis	73	17	9	11	9	8	6	60	–
Erysipelas	2	–	1	–	–	1	–	2	–
Impetigo	6	–	–	1	–	1	1	3	–
Infection skin	59	7	8	10	8	6	7	46	–
Paronychia	17	2	3	1	1	3	5	15	–
Lice	1	–	–	–	–	–	–	–	–
Lichen planus	1	–	1	–	–	–	–	1	–
Lupus discoid	1	–	–	–	1	–	–	1	–

^a patient-months of observation.

Analysis of Events During the Study/Events While on Drug

Table 10.3 shows the first page of a table that summarizes all reports received throughout a typical PEM study, whether or

not the patient was still on the drug. Denominators are given (in terms of patient-months of observation) for each month of the study, and for each of the 1700 or so events in the DSRU dictionary. The number of events reported is shown for each month

Table 10.4. Events reported on green forms during treatment with meloxicam (summarizing all reports received throughout a typical PEM study, restricted to events reported between the date of starting and stopping the drug being monitored)

Event	Total	Mth 1	Mth 2	Mth 3	Mth 4	Mth 5	Mth 6	Mths 1–6	Not known
Denominator total ^a	74 948	15 382	10 812	9497	8676	8036	7560		
Denominator male	23 711	4895	3384	2979	2723	2528	2377		
Denominator female	50 185	10 245	7282	6392	5840	5403	5084		
<i>Skin</i>									
Acne	10	1	1	–	3	1	1	7	–
Acne	5	1	–	–	1	–	1	3	–
Acne rosacea	5	–	1	–	2	1	–	4	–
Alopecia	1	–	–	–	1	–	–	1	–
Cyst sebaceous	9	3	3	1	1	–	1	9	–
Dermatitis	27	8	3	2	2	2	1	18	–
Dermatitis contact	3	2	–	–	–	1	–	3	–
Dry skin	17	5	3	–	3	1	1	13	–
Eczema	59	11	11	8	8	3	2	43	1
Eczema	44	7	8	6	5	3	2	31	1
Eczema atopic	1	1	–	–	–	–	–	1	–
Intertrigo	11	2	3	1	3	–	–	9	–
Pompholyx	3	1	–	1	–	–	–	2	–
Eczema varicose	6	2	–	1	1	–	1	5	–
Eruption bullous	4	–	1	1	1	–	–	3	–
Blister	4	–	1	1	1	–	–	3	–
Pemphigoid	–	–	–	–	–	–	–	–	–
Erythema	2	1	1	–	–	–	–	2	–
Erythema multiforme	1	–	1	–	–	–	–	1	–
Folliculitis	4	–	–	1	2	–	1	4	–
Granuloma	1	1	–	–	–	–	–	1	–
Granulomatosis	–	–	–	–	–	–	–	–	–
Haematoma nail	1	–	–	–	–	–	1	1	–
Hair loss	5	1	1	1	–	–	–	3	–
Herpes simplex, skin	1	–	1	–	–	–	–	1	–
Herpes zoster	28	7	3	4	5	6	1	26	–
Hyperkeratosis	3	–	1	1	–	1	–	3	–
Hyperkeratosis	2	–	1	–	–	1	–	2	–
Pityriasis	1	–	–	1	–	–	–	1	–
Infection skin, unspecified/local bacterial	124	25	13	16	16	19	13	102	–
Abscess skin	17	2	–	1	5	6	2	16	–
Cellulitis	52	16	5	9	5	6	4	45	–
Erysipelas	2	–	1	–	–	1	–	2	–
Impetigo	3	–	–	1	–	1	–	2	–
Infection skin	38	5	5	5	5	3	4	27	–
Paronychia	12	2	2	–	1	2	3	10	–
Lice	–	–	–	–	–	–	–	–	–
Lichen planus	1	–	1	–	–	–	–	1	–
Lupus discoid	–	–	–	–	–	–	–	–	–

^a Patient-months of observation.

of the study. Table 10.4 provides similar data but is restricted to events reported between the date of starting and stopping the drug being monitored. Each of these tables shows events grouped into system–organ class and displayed as higher and lower terms where the dictionary has been divided in this way. Each table also shows the total number of reports for each event, the total over the first six months of observation, and the number of events where the date of event was unknown. Comparison of these two tables (and a third table listing off-drug events of unknown date) indicates the number of reports for each event when the patients were not receiving the drug being monitored. This allows on-drug/off-drug comparisons (although the period after the drug being monitored has been withdrawn may be a period in which some other, unknown, drug was being given in individual patients).

These tables can generate signals: the total for an event may be unusually high and this can be confirmed or refuted by comparisons across the database of all 88 drugs that have been studied to date, or by comparison with drugs of the same therapeutic group with the same indication for use. The trend of reports over the months of observation may be informative: Type A side effects (pharmacologically related) 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). Again, formal trend analysis can be used to explore, on a comparative basis, such apparent signals. An example is weight gain with the atypical antipsychotic olanzapine.

Ranking of Incidence Density and Reasons for Withdrawal

The incidence density (ID) for a given time period t is calculated, for each event in the dictionary, in the usual way:

$$ID_t = \frac{N_t}{D_t} \times 1000$$

where N_t is the number of reports of the event during treatment for period t , D_t is the number of patient-months of treatment for period t , and the results are given in terms of 1000 patient-months of exposure. These results are then ranked in order of the estimate of ID_1 (the incidence density for the event in question in the first month of exposure). The incidence densities in the second to sixth months of treatment are also routinely calculated (ID_2).

Table 10.5 shows the first page of such a report of ranked incidence densities from a typical PEM study. For

each event, the table presents the value of ID_1 minus ID_2 and the 99% confidence intervals around this difference. This difference can itself generate signals, which require confirmation or refutation by further evaluation or another study. The basis for this is that most pharmacologically-related ADRs occur soon after initial exposure. However, for ADRs with long latency the comparison can be reversed, e.g., comparing the ID in month 6 with the ID in months 1–5.

The ranked reasons for withdrawal 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, and an example of this is given in Table 10.6; other examples have been published elsewhere.

Comparison of Event Rates and Adjusted Rates

Rate comparisons can be helpful in exploring apparent associations. An example occurred when looking at the gastrointestinal events of celecoxib (a COX-2 inhibitor) compared with the NSAID meloxicam.

Analysis showed that the adjusted rate ratio of symptomatic upper gastrointestinal events or complicated upper gastrointestinal conditions (perforations/bleeding) for rofecoxib compared with meloxicam were 0.77 (95% CI 0.69, 0.85) and 0.56 (95% CI 0.32, 0.96), respectively.

Examples of other signals and comparisons that have been explored include deaths from cardiac arrhythmias and suicide with atypical antipsychotics, sedation with non-sedating antihistamines, and bleeding with SSRIs.

Automated Signal Generation

The DSRU is exploring the application of automated signal generation as a possible additional tool in PEM. Feasibility studies apply comparisons of incidence rate ratios (IRRs). The exploratory work included confirming historical signals, e.g., Stevens–Johnson syndrome with the anti-epileptic product lamotrigine, and new signals such as exacerbation of colitis with rofecoxib. There are a number of methodological issues which need to be further examined with automated signal generation such as the selection of comparator(s) and the level of dictionary terms used, i.e., higher- or lower-level terms, because both factors may influence whether a signal is generated or its strength. However, with refinement, automated signal generation is likely to prove useful in spontaneous reporting, clinical trials, and pharmacoepidemiologic studies.

Table 10.5. Incidence densities (IDs) ranked for meloxicam in order of ID_1 per 1000 patient-months

Event	N_1	N_2	ID_1	ID_2	$ID_1 - ID_2$	99% CI		N_A	ID_A
						min	max		
1 Condition improved	903	1015	58.7	22.8	35.9	30.6	41.3	2067	27.6
2 Dyspepsia	435	379	28.3	8.5	19.8	16.1	23.5	903	12.0
3 Respiratory tract infection	214	387	13.9	8.7	5.2	2.5	7.9	675	9.0
4 Nausea, vomiting	189	136	12.3	3.1	9.2	6.8	11.6	351	4.7
5 Pain abdomen	146	163	9.5	3.7	5.8	3.7	8.0	357	4.8
6 Diarrhea	118	110	7.7	2.5	5.2	3.3	7.1	251	3.3
7 Dose increased	106	206	6.9	4.6	2.3	0.4	4.2	353	4.7
8 Headache, migraine	81	82	5.3	1.8	3.4	1.8	5.0	192	2.6
9 Minor surgery	75	100	4.9	2.2	2.6	1.1	4.2	198	2.6
10 Hospital referrals no admission	74	144	4.8	3.2	1.6	0.0	3.2	270	3.6
11 Edema	74	96	4.8	2.2	2.7	1.1	4.2	195	2.6
12 Dizziness	68	70	4.4	1.6	2.9	1.4	4.3	152	2.0
13 Gastrointestinal unspecified	66	64	4.3	1.4	2.9	1.4	4.3	144	1.9
14 Pain joint	57	143	3.7	3.2	0.5	-0.9	1.9	236	3.1
15 Malaise, lassitude	56	81	3.6	1.8	1.8	0.5	3.2	153	2.0
16 Constipation	56	57	3.6	1.3	2.4	1.0	3.7	119	1.6
17 Hospital referral paramedical	51	55	3.3	1.2	2.1	0.8	3.4	111	1.5
18 Cough	48	73	3.1	1.6	1.5	0.2	2.7	140	1.9
19 Asthma, wheezing	48	46	3.1	1.0	2.1	0.9	3.3	107	1.4
20 Rash	46	90	3.0	2.0	1.0	-0.03	2.2	151	2.0
21 Nonsurgical admissions	45	59	2.9	1.3	1.6	0.4	2.8	127	1.7
22 Noncompliance	42	61	2.7	1.4	1.4	0.2	2.5	113	1.5
23 Hypertension	42	57	2.7	1.3	1.5	0.3	2.6	119	1.6
24 Intolerance	42	21	2.7	0.5	2.3	1.1	3.4	71	0.9
25 Urinary tract infection	39	99	2.5	2.2	0.3	-0.9	1.5	163	2.2
26 Depression	38	76	2.5	1.7	0.8	-0.4	1.9	139	1.9
27 Pain back	29	87	1.9	2.0	-0.1	-1.1	1.0	142	1.9
28 Dyspnoea	29	38	1.9	0.9	1.0	0.1	2.0	73	1.0
29 Fall	26	77	1.7	1.7	0.0	-1.0	1.0	115	1.5
30 Infection skin, unspecified	25	77	1.6	1.7	-0.1	-1.1	0.9	124	1.7
31 Malignancies	25	33	1.6	0.7	0.9	0.0	1.8	68	0.9
32 Unspecified side effects	25	13	1.6	0.3	1.3	0.5	2.2	42	0.6
33 Pain in chest, tight chest	24	60	1.6	1.3	0.2	-0.7	1.2	94	1.3
34 Pain	24	48	1.6	1.1	0.5	-0.4	1.4	84	1.1
35 Ischemic heart disease	24	43	1.6	1.0	0.6	-0.3	1.5	78	1.0
36 Cardiac failure	24	29	1.6	0.7	0.9	0.0	1.8	58	0.8
37 Pruritus	22	50	1.4	1.1	0.3	-0.6	1.2	84	1.1
38 Ulcer, mouth	21	43	1.4	1.0	0.4	-0.5	1.3	71	0.9
39 Hematological tests	20	39	1.3	0.9	0.4	-0.4	1.3	65	0.9
40 Palpitation	20	22	1.3	0.5	0.8	0.0	1.6	46	0.6
41 Orthopaedic surgery	19	95	1.2	2.1	-0.9	-1.8	0.0	139	1.9
42 Osteoarthritis	17	53	1.1	1.2	-0.1	-0.9	0.7	83	1.1
43 Micturition disorder	16	39	1.0	0.9	0.2	-0.6	0.9	61	0.8
44 Anxiety	16	34	1.0	0.8	0.3	-0.5	1.0	58	0.8

Long Latency Adverse Reactions

There is interest in reactions that emerge only on prolonged treatment and may be missed in the premarketing trials, many of which are frequently of short duration. An example occurred in the PEM study of finasteride, a product used

for the treatment of benign prostatic hypertrophy, when it was shown that reports of impotence/ejaculatory failure and decreased libido were received in relation to the first and all subsequent months of treatment, but reports of gynecostasia were only rarely received before the fifth month of therapy. A further important example has occurred in

Table 10.6. Most frequently reported events

Reason	Number
Not effective	2989
Condition improved	1834
Dyspepsia	539
Nausea, vomiting	209
Pain abdomen	171
Noncompliance	117
Gastrointestinal unspecified	104
Diarrhea	103
Orthopedic surgery	87
Effective	81
Headache, migraine	72
Hospital referrals no admission	69
Rash	64
Dizziness	60
Intolerance	59
Malaise, lassitude	58
Patient request	50
Edema	49
Minor surgery	45
Hospital referral paramedical	38
Nonsurgical admissions	36
Pain	33
Unspecified side effects	33
Constipation	31
Asthma, wheezing	30
Ulcer, mouth	23
Indication for meloxicam changed	21
Pruritus	20
Nonformulary	18
Dyspnea	16
Hemorrhage gastrointestinal, unspecified	16
Surgery, unspecified	16
Tinnitus	16
Distension, abdominal	12
Drowsiness, sedation	12
Hemorrhage gastrointestinal upper	12
Pain in chest, tight chest	12
Anemia	11
Pain, joint	11
Hemorrhage rectal	10

relation to visual field defects in patients receiving long-term treatment with vigabatrin. The initial PEM study showed three cases of bilateral, irreversible peripheral field defects, whereas no similar reports occurred with other anti-epileptic drugs or in any of the other drugs already monitored by PEM. A follow-up exploration with a repeat questionnaire, sent to the doctors whose patients had received vigabatrin for over six months, has shown that the incidence of this serious event is much higher and that many of the relevant patients have objective evidence of visual field defects.

Comparison with External Data

With 88 completed PEM studies to date, there are increasing opportunities to conduct comparisons among PEM studies. However, it is not always possible to identify suitable comparators. Therefore, external comparators are sought where necessary and appropriate. For example, there were concerns about cardiovascular safety when sildenafil (the first PDE5 inhibitor marketed for erectile dysfunction) was launched in the UK in 1998. Mortality from ischemic heart disease in users of sildenafil in the PEM study was compared with external epidemiologic data for men in England. The standardized mortality ratio (SMR) for deaths reported to have been caused by ischemic heart disease was not higher for sildenafil users: SMR 69.9 (95% CI, 42.7–108.0). 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. Obviously there is higher potential for bias when using external comparators than comparisons undertaken between PEM studies; results of external comparisons must be considered very carefully.

Outcomes of Pregnancy

There is interest in determining the proportion and nature of congenital anomalies in babies born to women exposed to newly marketed drugs during the first trimester. PEM studies have shown that from 831 such pregnancies 557 infants were born, of whom 14 (2.5%) had congenital anomalies. Projects are under way to compare pregnancy outcomes following drug exposure between PEM studies or between 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.

Studies to Examine Hypotheses Generated by Other Methods

In addition to examining signals generated in PEM, the database provides a resource that is being used increasingly to evaluate signals and hypotheses generated by other methods. An example of such studies is the comparison of mortality and rates of cardiac arrhythmias with atypical antipsychotic drugs.

STUDIES OF BACKGROUND EFFECTS AND DISEASES

Background Effects

The PEM database allows the study of diseases as well as drugs. An example includes a study of the prevalence of Churg–Strauss syndrome and related conditions in patients with asthma. The study defined the period prevalence rate for this condition, 6.8 (95% CI, 1.8–17.3) per million patient-year of observation, and demonstrated a much higher period prevalence rate in patients receiving asthma medications compared to other PEM cohorts. In another study, the PEM database was used to define age- and gender-specific asthma deaths in patients using long-acting β_2 -agonists.

Study of the database also shows some of the characteristics of ADR reporting. Doctors are asked to note on the green form if they have previously reported an event spontaneously as an ADR (in a patient being monitored by PEM). Two studies compared events that were considered as ADRs by doctors reported in PEM, with spontaneous reports sent by the same doctors to the regulatory authority. The studies showed that reporting of suspected spontaneous ADRs to the UK regulatory authority was 9% (95% CI, 8.0–10.0) and 9% (95% CI, 8.0–9.8), respectively. In the more recent study, published in 2001, it was shown that, of 4211 ADRs reported on the PEM green form questionnaires, only 376 (9.0%) had also been reported on yellow cards to the CSM. It is of interest that a higher proportion of serious reactions were reported to the CSM by doctors, which suggests that doctors use the spontaneous adverse reaction reporting scheme more energetically when reporting those serious reactions that worry them most.

It is possible to study in PEM 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.

THE FUTURE

In the future, PEM aims to utilize improvements in information technology, application of additional study designs such as nested case–control studies, and the application of new biological developments such as pharmacogenetics to enhance the PEM process. Modification of the PEM method is sometimes necessary to examine specific drug safety questions. In addition, it is possible to modify the PEM

process to examine questions related to risk management of marketed medicinal products.

NESTED CASE–CONTROL STUDIES

PEM cohorts provide opportunities to conduct nested case–control studies, for example, for patients who develop selected ADRs and matched patients who receive the same drug without developing ADRs. A nested case–control study is planned to study patients who were reported to have had ischemic cardiac events in the cohort of users of the PDE5 inhibitor tadalafil and matched controls, to examine the risk factors for events such as hypertension, smoking, etc. There are plans to broaden the scope for the application of nested case–control studies to PEM.

PHARMACOGENETICS

There is increasing interest in understanding the role of pharmacogenetics in the efficacy and safety of medicines (see Chapter 18). Given the interest in understanding the roles of polymorphic genotypes of receptors, protein carriers, and metabolizing enzymes of drugs, there are many opportunities in PEM to study the genotypes of patients who develop selected ADRs compared to patients who do not develop such ADRs. Moreover, there are opportunities to study the genotypes of patients who do not respond to some drugs. Nested case–control pharmacogenetic studies of both types are under way in PEM.

MODIFIED PEM STUDIES

In some cases, it is considered necessary to modify PEM methodology to answer specific safety question(s) regarding the safety of a particular product. A study is under way to examine specific eye events (discoloration of the iris and lengthening of eye lashes) that have been reported following the use of an ophthalmic product used for the treatment of glaucoma. The length of the PEM follow-up and details of the outcome questionnaires were modified in order to answer the specific research questions.

RISK MANAGEMENT

Risk management is attracting immense interest in pharmacovigilance (see Chapter 27). 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 does not only contribute 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. Two studies are under way on two new antidiabetic agents, rosiglitazone and pioglitazone, where detailed questionnaires are 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, as well as their outcomes. Another study is underway to monitor the introduction of carvedilol for the treatment of cardiac failure. 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 modified PEM study is 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, etc. The design includes sending an eligibility questionnaire followed by a maximum of three detailed questionnaires over a period of up to two years.

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 methods, PEM is evolving, aiming to reduce its weaknesses and enhance its strengths. New methodological 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.

ACKNOWLEDGMENTS

PEM is a team effort and I am only one member of a large team. The DSRU is most grateful to the thousands

of doctors across England who provide the Unit, free of charge, with the safety information which makes its public health work possible. The Unit is equally grateful to the PPD; PEM would not be possible without their immense support.

I am most grateful to previous and current staff of the DSRU; this chapter is based on their work! Special gratitude goes to Professor Ron Mann for allowing me to use material from the previous edition, which he wrote, and to Georgina Spragg and Lesley Flowers, who helped in locating research material and typing the manuscript.

Key Points

- While premarketing clinical trials identify many aspects of drug safety, the full safety picture can only be identified when the drug is used in large number of users in "real world" clinical settings.
- Prescription-Event Monitoring (PEM) systematically monitors the safety of new medicines in cohorts of 10 000–15 000 patients and is complementary to spontaneous reporting of suspected ADRs.
- PEM is a method for both hypothesis generation and testing.
- PEM offers opportunities for better quantification of ADRs and identifies and characterizes some ADRs which were unrecognized during premarketing development and are not possible to quantify by spontaneous reporting.
- Clinical and quantitative analyses are complementary and utilized in PEM for the understanding of drug safety.
- More than 90 PEM studies have been completed and there are now opportunities to compare drugs in the same therapeutic class.
- Recently it has been possible to modify the PEM methodology to conduct studies to explore specific safety questions and studies which compare events before and after exposure to medicines.

SUGGESTED FURTHER READINGS

- Biswas PN, Wilton LV, Pearce GL, Freemantle S, Shakir SA. The pharmacovigilance of olanzapine: results of a post-marketing surveillance study on 8858 patients in England. *J Psychopharmacol* 2001; **15**: 265–71.
- Finney DJ. The design and logic of a monitor of drug use. *J Chron Dis* 1965; **18**: 77–98.
- Heeley E, Riley J, Layton D, Wilton LV, Shakir S. Prescription-event monitoring and reporting of adverse drug reactions. *Lancet* 2001; **356**: 1872–3.

- Inman WHW, Weber JCP. Post-marketing surveillance in the general population. In: Inman WHW, Gill EP, eds. *Monitoring for Drug Safety*, 2nd edn. Lancaster: MTP, 1985; p. 13.
- 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**(1): 98–102.
- Wilton LV, Stephens MDB, Mann RD. Visual field defect associated with vigabatrin: observational cohort study. *BMJ* 1999; **319**: 1165–66.
- Wilton LV, Heeley EL, Pickering RM, Shakir SAW. Comparative study of mortality rates and cardiac dysrhythmias in post-marketing surveillance studies of sertindole and two other atypical antipsychotic drugs, risperidone and olanzapine. *J Psychopharmacol* 2001; **15**: 120–6.

11

Overview of Automated Databases in Pharmacoepidemiology

Edited by:

BRIAN L. STROM

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

INTRODUCTION

Once hypotheses are generated, usually from spontaneous reporting systems (see Chapters 7 and 8), 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 3). 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 3). 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 two decades have seen a growing use of

computerized databases containing medical care data, so-called “automated databases,” as potential data sources for pharmacoepidemiology 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 normally 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 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.

Of course, no single database is ideal. In the current chapter, we introduce these resources, presenting some of the general principles that apply to them all. In Chapter 12, we introduce 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 DATABASES

Claims data 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.

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, e.g., by Medicaid (see Chapter 12). 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, i.e., 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 15).

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

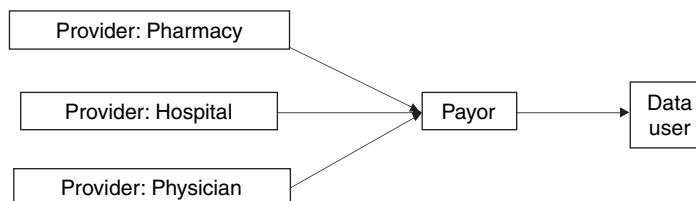


Figure 11.1. Sources of claims data.

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.

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 General Practice Research Database, described in Chapter 12.

Medical record databases have unique advantages. Important 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 pharmacoepidemiology study using these databases, there is no need to validate the data against the actual medical record, since the physician-made diagnosis is already recorded. 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, i.e., 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.

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. It is less problematic for inpatient diagnoses and for medical record databases (see Chapters 12 and 15).

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 to physician records, if they 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 such data on variables which may be important potential confounders that must be controlled for. Alternatively, one can perform sensitivity analyses to determine how one's results might change under the assumption of the presence of uncontrolled confounders. However, this latter approach relies on making various assumptions about the confounders, and, if the results are sensitive to uncontrolled confounding, one is left with an inconclusive study.

Computerized databases also do not typically include information on medications obtained without a prescription or outside of the particular insurance carrier's prescription plan. Information on over-the-counter medication use (including dietary supplements) can be particularly important if it is a large component of the drug exposure of interest (e.g., nonsteroidal anti-inflammatory drugs) or an important confounder or effect modifier of the drug-outcome exposure being investigated (e.g., the effect of aspirin on the association between nonsteroidal anti-inflammatory drugs and gastrointestinal bleeding). Similarly, even the use of a prescription drug may not be captured if the drug is obtained outside of the prescription plan (e.g., through sharing with a spouse; obtaining medications from the internet; self-payment for medications, especially if inexpensive and

below a deductible; or, rarely, use of alternative prescription plans).

A major other disadvantage of claims-based data is the instability of the population due to job changes, employers' changes of health plans, and changes in coverage for specific employees and their family members. The opportunity for longitudinal analyses is thereby hindered by the continual enrollment and disenrollment of plan members. However, strategies can be adopted for selecting stable populations within a specific database, and for addressing compliance, for example, by examining patterns of refills for chronically used medications.

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.

Finally, some results from studies that utilize these databases may not be generalizable, e.g., on health care utilization. This is especially relevant for databases created by data from a population that is atypical in some way, e.g., US Medicaid data (see Chapter 12).

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 the need for 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, since the population might be 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 patient, such as cigarette smoking, occupation, menarche, menopause, etc.
7. important medication exposure information that is not available, particularly over-the-counter medications.

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. 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, diagnoses under study are chosen carefully, and medical records can be obtained to validate the diagnoses. In Chapter 12, we 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 in 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 13) 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 pharmacoepidemiology studies.

Key Points

- The past two decades have seen a growing use of computerized databases containing medical care data, so-called “automated databases,” as potential data sources for pharmacoepidemiology studies.
- Claims data arise from a person's use of the health care system, and the submission of claims to insurance companies for payment. While claims data provide some of the best data on drug exposure in pharmacoepidemiology, the quality of disease data in these databases can be more problematic.

- Medical record databases are a more recent development, arising out of the increasing use of computerization in medical care. 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. However, the completeness of the data from other physicians and sites of care is uncertain.

SUGGESTED FURTHER READINGS

- Ray WA, Griffin MR. Use of Medicaid data for pharmacoepidemiology. *Am J Epidemiol* 1989; **129**: 837–49.
- Strom BL, Carson JL. Automated data bases used for pharmacoepidemiology research. *Clin Pharmacol Ther* 1989; **46**: 390–4.
- Strom BL, Carson JL. Use of automated databases for pharmacoepidemiology research. *Epidemiol Rev* 1990; **12**: 87–107.

12

Examples of Automated Databases

The following individuals contributed to editing sections of this chapter:

ANDY STERGACHIS,¹ KATHLEEN W. SAUNDERS,² ROBERT L. DAVIS,³
STEPHEN E. KIMMEL,⁴ RITA SCHINNAR,⁴ K. ARNOLD CHAN,⁵ DEBORAH SHATIN,⁶
NIGEL S.B. RAWSON,⁶ SEAN HENNESSY,⁷ WINANNE DOWNEY,⁸ MARYROSE STANG,⁸
PATRICIA BECK,⁸ WILLIAM OSEI,⁸ HUBERT G. LEUFKENS,⁹ THOMAS M.
MACDONALD,¹⁰ and JOEL M. GELFAND¹¹

¹Departments of Epidemiology and Pharmacy, University of Washington, Seattle, Washington, USA; ²Center for Health Studies, Group Health Cooperative, Seattle, Washington, USA; ³Center for Health Studies, Group Health Cooperative, and Departments of Pediatrics and Epidemiology, University of Washington, Seattle, Washington, USA; *Immunization Safety Office, Office of the Chief Science Officer, Centers for Disease Control and Prevention, Atlanta, Georgia, USA;

⁴University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; ⁵Harvard Medical School, Boston, Massachusetts, USA; ⁶Center for Health Care Policy and Evaluation, UnitedHealth Group, Minneapolis, Minnesota, USA;

⁷University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; ⁸Population Health Branch, Saskatchewan Health, Regina, Saskatchewan, Canada; ⁹Department of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute for Pharmaceutical Sciences (UIPS), The Netherlands; ¹⁰Medicines Monitoring Unit, University of Dundee, Ninewells Hospital & Medical School, Dundee, UK; ¹¹Department of Dermatology and Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

This chapter describes nine existing medical databases that have been useful in the conduct of pharmacoepidemiologic research, including several domestic US databases as well as some notable international ones. These databases run the gamut from national to regional and from private to national governmental sponsorship. In presenting these

various databases, we will follow the same format, outlining the description, strengths, and weaknesses of each. In order, the databases covered include: Group Health Cooperative, Kaiser Permanente Medical Care Program, HMO Research Network, UnitedHealth Group, Medicaid Databases, Health Services Databases in Saskatchewan,

* Current post

Automated Pharmacy Record Linkage in The Netherlands, Tayside Medicines Monitoring Unit (MEMO), and UK General Practice Research Database (GPRD).

GROUP HEALTH COOPERATIVE

Traditional group and staff model health maintenance organizations (HMOs) employ a defined set of providers to deliver comprehensive health care services to a defined population of patients for a fixed, prepaid annual fee. HMO enrollees typically receive their health care services within these integrated systems through a uniform benefit package. Care is usually provided within a defined geographic area that allows for large comparable groups of subjects for public health research. The majority of employed Americans are now covered by some type of managed health care plan.

There has been longstanding interest in the use of data from HMOs to study the effects of marketed drugs. There are several advantages to conducting research in an HMO setting: because every HMO has an identifiable population base, it is possible to determine denominators for epidemiologic research, enabling investigators to calculate incidence and prevalence rates; also, managed care organizations often use pharmacy benefits management (PBM) companies to perform some or all of the management of prescription drug benefits, thus providing data on medication prescriptions. Other key features of HMOs relevant to the conduct of postmarketing drug surveillance include the availability of: (i) a relatively stable population base; (ii) accessible and complete medical records for each enrollee; and (iii) in many instances, computerized databases. In general, these automated files contain information recorded during the routine delivery of health services. At Group Health Cooperative (GHC) such data have been used extensively to evaluate drug usage and the adverse and beneficial effects of marketed drugs and medical procedures.

DESCRIPTION

GHC, a nonprofit consumer-directed HMO established in 1947, currently provides health care on a prepaid basis to approximately 562 000 persons in Washington state. About 75% of these enrollees are part of the “staff model”—that is, they receive outpatient care at GHC facilities, with the exception of specific services not provided by GHC providers (e.g., temporomandibular care). Approximately 25% of enrollees deviate from the staff model in that they receive care from non-GHC provider networks located in geographic areas not served by GHC medical centers.

Approximately 10% of Washington residents were enrolled in GHC in 2003. The majority of GHC enrollees receive health benefits through their place of employment (i.e., group enrollees). As of September 2003, GHC had arrangements for providing services to approximately 58 500 Medicare patients, 30 000 Medicaid patients, and 18 000 patients in a state-subsidized program that provides medical insurance to low income, uninsured residents.

Historically, GHC offered comprehensive health care coverage for care. However, this approach to comprehensive coverage has changed over the last decade. Beginning in 1993, Medicare enrollees new to GHC did not receive drug coverage, but could purchase prescription drugs from GHC pharmacies at prices competitive with the rest of the community. Even among others with comprehensive coverage, nearly all benefit plans required small copayments for services.

Compared to other Seattle–Tacoma–Bremerton area residents, GHC enrollees are slightly better educated but similar in age, gender, and racial/ethnic composition. GHC enrollees have similar median income, but there is less representation within the highest extreme of income distribution. Differences noted between the GHC population and the US primarily reflect differences between the demographic composition of Seattle–Tacoma–Bremerton and the US population as a whole.

GHC’s automated and manual databases serve as major resources for many epidemiologic studies, in part because individual records can be linked through time and across data sets by the unique consumer number assigned to each enrollee. Once assigned, the consumer number remains with an enrollee, even if the individual disenrolls and rejoins GHC at a later date. Files are routinely updated using data from clinical and administrative computer systems. Every file contains the unique patient identifier common to all of the data sets. Physician identifiers are also unique across all files. A brief description of each of GHC’s data files follows.

Enrollment

GHC maintains various enrollment and demographic files. Current enrollment files contain records for every person presently enrolled. They contain person-based information on selected patient characteristics such as patient consumer number, subscriber number (used to aggregate family members on the same contract), date of birth, sex, primary care provider, plan, assigned clinic, patient address, and telephone number. (Note that information on race, years of education, and income is not routinely collected, but was obtained through special surveys from random samples

of enrollees.) This file is often used to select probability samples of current GHC enrollees. GHC also maintains historical enrollment files.

Pharmacy

The pharmacy file includes data on each prescription dispensed at GHC-owned outpatient pharmacies since March 1977. A computerized record is created for every medication at the time the prescription is filled. Each record contains selected information about the patient, the prescription, and the prescriber, including: patient data, consumer number, birth date, sex, group coverage, copay status, drug data, drug number, therapeutic class, drug form and strength, date dispensed, quantity dispensed, cost to GHC, refill indicator, number of days' supply dispensed, prescriber or pharmacy data, prescribing physician, and pharmacy location. An additional variable, days' supply, was added to the pharmacy database in 1996.

Hospital

The hospitalization databases contain records of every discharge, including newborn and stillborn infants, from GHC-owned hospitals since January 1972. The information in the hospitalization database includes patient characteristics, diagnoses, procedures, diagnostic related group (DRG), and discharge disposition. Virtually all hospitalizations not captured in GHC's hospital information system are included in GHC's outside claims files.

Laboratory and Radiology

Automated laboratory data are available from January 1986. The online laboratory system interconnects all GHC laboratories, including inpatient settings, and contains patient-specific information on all laboratory tests. Specific variables contained in the research file include name of the test ordered, date ordered, specimen source, results, and date of the results. The radiology system, established in 1986, contains records of all radiographic studies performed at GHC facilities, including CT and MRI scans.

Outpatient Visits

The GHC outpatient registration system was initiated in mid-1984 and includes selected information about each outpatient visit. The database includes date of visit, provider seen, provider's specialty, and location of care. Beginning in 1991, and fully operational in 1992, diagnosis and

procedure data have been incorporated in the registration database.

Cancer Surveillance System and Other Registries

Since 1974, GHC has participated in the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program. As part of this program, the Center for Health Studies periodically receives data files with information on all newly diagnosed cancers among GHC enrollees from the Cancer Surveillance System (CSS) at the Fred Hutchinson Cancer Research Center, one of 13 SEER population-based registries in the US. The database contains information for each newly diagnosed cancer case, including patient demographics, anatomical site, stage at diagnosis, and vital status at follow-up, which is ongoing for all surviving cases in the register. GHC has also developed several electronic disease-specific registries that identify relevant populations of patients, including registries of patients with diabetes, depression, and heart disease.

Cause of Death

Using data from Washington vital statistics files, a file of deaths among GHC enrollees between the years 1977 to the present has been developed. The file was produced through record linkages between Washington's death certificate database and the GHC membership files.

Community Health Services System

Computerized information from visiting nurse, hospice, respite, and geriatric nurse practitioner programs, a nursing home rounding system, and numerous other community-based service programs includes type of provider, type of procedure, diagnosis, location of service, and price and billing information.

Utilization Management/Cost Management Information System

GHC developed a Utilization Management/Cost Management Information System (UM/CMIS) in 1989. This system uses information from other systems (e.g., registration, pharmacy) to assign direct and indirect costs to individual encounters based on the units of service utilized. Thus, UM/CMIS data can be used to estimate an individual's total health care costs, as well as the costs of individual components of care, such as primary care visits, pharmacy prescriptions, mental health services, and inpatient stays.

Immunizations Database

The immunizations database was initially developed as part of the Vaccine Safety Datalink (VSD), a collaborative study funded by the Centers for Disease Control including GHC and seven other HMOs nationwide. This database contains immunization records from February 1991 to the present for GHC enrollees 0–6 years of age, while immunizations for GHC members of all ages have been included since 1995. GHC also regularly receives immunization data from the state for all GHC-enrolled children, including historical data prior to their status as enrollees (if applicable), thereby preventing a loss of immunization information when children change health plans or providers.

Claims Database

The claims databases contain information on health care services that GHC purchases from non-GHC providers. These claims databases have become even more critical as traditional HMOs have shifted to more mixed models of care, because they provide a full accounting of an individual's medical care.

Other Databases

Various manually maintained outpatient and inpatient medical records, clinical logs, and registries are often employed in research. Many research studies include direct patient contact through in-person interviews, surveys, or clinical examinations and tests.

STRENGTHS

The advantages of the GHC databases derive from their size, the accessibility and quality of their data, and the feasibility of linking patients to a primary care provider in a staff model HMO. GHC serves a moderately large population, currently over 562 000 persons. The GHC population is well defined in terms of age, sex, geographic location, and duration of GHC membership. Turnover in membership is estimated to be 15% per year. This relative stability facilitates long-term follow-up.

An obvious advantage to the GHC setting for research is the extensive use of computer technology. Because the data are collected in an ongoing manner as a by-product of health care delivery, there is no requirement for individual patient contact. (See Case Example 12.1.) Thus, the high cost and potential for recall bias that would otherwise be associated with primary data collection are minimized or avoided. A single patient identification number permits following the records of individuals across time. However,

another advantage is that GHC has mechanisms for contacting patients directly for primary data collection. Examples include interview-based case-control studies, health surveillance activities, and clinical research studies. Manual medical records and registries are also available for use. Another advantage is the longevity of the plan. In existence for more than 50 years, GHC has a subset of enrollees whose tenure in the cooperative spans decades. This longevity facilitates studies that require very long-term follow-up.

Another benefit is the ability to study rare events. Since specific exposures such as medications or vaccinations are easily identifiable by automated data routinely collected for billing purposes, studies of rare events following these exposures can often be performed in a more efficient and less biased fashion than otherwise possible. Similarly, using automated pharmaceutical files to track medication usage, studies can be done of rare adverse events following use of specific medications or among subjects with specific conditions.

WEAKNESSES

Logistic and Operational

Despite the large size of the GHC databases, most marketed drugs are used by a relatively small proportion of a population. Thus, the GHC databases still may be too small to detect associations between drug exposures and rare outcomes. The detection of rare adverse events requires combining data from multiple health care delivery systems, as has been accomplished by the Centers for Education and Research on Therapeutics (CERTs, see Chapter 6).

The data presently available in automated files do not include some potentially important confounding variables. The lack of information on confounding factors such as race, smoking, and alcohol consumption may lead to challenges in the study and interpretation of drug-associated health effects. Analytic studies must consider alternative methods of obtaining the necessary information on confounding factors, such as medical record abstraction and patient interview. It should be noted that GHC's new Clinical Information System, EpicCare, may help solve some of these problems. Also, automated information on inpatient drugs and outpatient diagnoses was not available until relatively recently, an important disadvantage for some retrospective studies.

Use of automated data to determine health outcomes is not always reliable without a review of medical records. However, as noted above, medical records are available for outcome validation.

The present competitive environment in the health care industry has led GHC to offer more mixed-model benefit plans (e.g., non-GHC provider networks and point-of-service

plans), which may impact the completeness of databases of health services utilization. For example, results of laboratory tests administered outside of GHC are not presently available in electronic form. Another threat to database completeness is the movement away from “one size fits all” comprehensive benefit packages. This has resulted in varying coverage arrangements for different groups. For example, from 1994 until the recent passage of Medicare Part D, GHC has not offered prescription drug coverage to new Medicare enrollees. While using automated data to assess current coverage arrangements is possible, albeit complicated, the data are not structured to facilitate retrospective inquiries into coverage. Thus, it is very difficult to track changes in coverage over time, which could have serious ramifications for health care utilization. It should be noted that out-of-plan use of prescription drugs has been the subject of several validity studies, and the outpatient pharmacy database has been found to be generally complete for prescription drugs.

In some populations, even when individuals are part of the “staff model” and have comprehensive coverage, they may choose to receive some of their care out-of-plan. For example, one study found that a large proportion of GHC adolescents used out-of-plan care, and were also more likely to have sexually transmitted diseases and other health problems than those who solely used in-plan services. Aside from jeopardizing continuity of care, out-of-plan use limits the completeness of automated databases. Increased competition may also lead to increased patient turnover in HMOs, resulting in decreases in follow-up time for cohort studies.

The GHC formulary limits the study of many newly marketed drugs, since GHC may decide not to add a new agent or may adopt a new drug only after it has been on the market for some time. GHC often maintains only one brand of a legend drug on the drug formulary at a time, thereby preventing investigations of many direct drug-to-drug comparisons of relative toxicity and effectiveness. If nonformulary drugs are commonly purchased outside the GHC pharmacy system, as may have been the case with drugs for impotence and some drugs for weight loss, there is the potential for an inaccurate determination of the prevalence of usage, or even of the risk of use. Furthermore, this situation (outside procurement of nonformulary drugs) prevents GHC from proactively contacting enrollees if new information emerges concerning the potential risks or dangers of medication.

Studies of some medications that are solely (or primarily) administered as over-the-counter (OTC) products are limited since such OTC data are not routinely captured by the GHC pharmacy database.

The elderly and the poor tend to be underrepresented in HMOs, leading to concerns about representativeness of studies. However, a Medicare managed care plan has been offered since 1985. GHC’s involvement in Healthy Options,

Washington’s Medicaid managed care program, resulted in a large increase in Medicaid enrollment between 1993 and 2003. However, due to financial considerations, the cooperative decided to reduce its Healthy Options enrollment beginning in 2004.

CASE EXAMPLE 12.1: **ASSESSING THE EFFECTS OF DRUG** **EXPOSURES ON PREGNANCY OUTCOMES**

Background

- Depression is common among women of childbearing age, and affected women are frequently prescribed antidepressant drugs.
- There is uncertainty about the safety of these drugs for offspring, when taken during pregnancy.
- The safety of antidepressant drugs used during pregnancy is an important public health issue.

Question

- What are the effects of prenatal antidepressant exposure on perinatal outcomes, congenital malformations, and early growth and development?

Approach

- An historical cohort study compared the perinatal outcomes, congenital malformations, and early growth and development of infants with and without prenatal exposure to antidepressants.
- The first step was to use hospital discharge records to identify all live births between January 1, 1986 and December 31, 1998. The newborn’s discharge record could be linked to that of the mother. To ensure that paper medical records and computerized data were available, the researchers required that mothers be enrolled at medical centers owned by GHC.
- The next step was to use the computerized pharmacy database to identify all tricyclic and serotonin reuptake inhibitor (SSRI) antidepressant prescription fills for the mothers during the 360 days prior to delivery. Women who had no antidepressant prescriptions in the 360 days preceding delivery were defined as “unexposed.” Mothers who had at least one antidepressant prescription during the 270 days before delivery were considered “exposed.” Those who fell somewhere in between were excluded from the analyses. Again, to ensure completeness of information on exposure (e.g., antidepressant use), the researchers required that

(Continued)

the mothers be enrolled continuously at GHC for the 360 days before delivery.

- Infants who, according to the above definition, were exposed to antidepressants were frequency matched to infants who were unexposed.
- Chart reviewers blinded to the infant's exposure status reviewed paper medical records for information pertaining to perinatal outcomes, congenital malformations, and developmental delay.

Results

- Infants exposed to tricyclic antidepressants ($n = 209$) or SSRIs ($n = 185$) during pregnancy were not at an increased risk for congenital malformations or developmental delay.
- Exposure to SSRIs during the third trimester of pregnancy was associated with lower Apgar scores.
- Infants exposed to SSRIs at any time during pregnancy were at an increased risk for premature delivery and lower birth weight than those infants not exposed to SSRIs.
- Tricyclic antidepressants did not increase the risk of premature delivery, lower birth weights, or lower Apgar scores.

Strengths

- Study infants were systematically selected from a defined population.
- Exposed and nonexposed subjects were matched for factors associated with gestational age, low birth weight, malformation, or developmental delay.
- Data collection and primary data analyses were completed without knowledge of exposure status.

Limitations

- Exposure represented drug dispensing records rather than actual ingestion of drugs.
- Outcomes information was based on recorded clinical data rather than specific examinations for malformation or developmental delay.
- Live births were sampled instead of pregnancies, resulting in the inability to assess the risk of spontaneous abortion.
- The 270-day period to defining pregnancy exposure to drugs may lead to some misclassification in situations of premature delivery.
- The study did not include information on possible exposures to antidepressant drugs used during breastfeeding or other postnatal exposures.

- The large number of comparisons performed in this analysis may affect marginally significant results.

Summary Points

- Infants exposed to tricyclic antidepressants ($n = 209$) or SSRIs ($n = 185$) during pregnancy were not at an increased risk for congenital malformations or developmental delay.
- SSRI use during pregnancy was associated with premature delivery, however the absolute risk was 10%.
- The authors concluded that women considering use of SSRIs during pregnancy may weigh "any greater risk of premature delivery against the risk of persistent or recurrent depression and the availability and acceptability of alternative treatments." This information may help women and their health care providers to make an informed decision on whether or not to use antidepressant drugs during pregnancy.
- Epidemiologic studies using medical databases can provide important information on the teratogenic effects of marketed drugs.

KAISER PERMANENTE MEDICAL CARE PROGRAM

The Kaiser Permanente (KP) Medical Care Program, with approximately 8.2 million subscribers nationally, is by far the largest and also one of the oldest prepaid, group model health care systems in the US. The KP program is divided administratively into eight regions, seven of which (Colorado, Georgia, Hawaii, Mid-Atlantic, Northern California, Northwest, and Southern California) have research departments that conduct public domain research, i.e., research funded and conducted with the understanding that results will be published and disseminated outside the organization. With approval from regional institutional review boards, researchers in each center access a host of administrative and clinical databases, paper medical records dating back as much as 50 years, and member populations through interviews, surveys, and direct clinical examinations. Within KP, each center is a distinct entity and each uses only the databases maintained within its region. Across regions, researchers are affiliated through KP's National Research Council. Most regional centers also participate along with other health maintenance organization (HMO)-based research units in the HMO Research Network, which sponsors various collaborative, multicenter projects, including pharmacoepidemiology studies.

Pharmacoepidemiology studies have been prominent for many years in the research portfolios of two research centers: the Division of Research in the Northern California region and the Center for Health Research in the Northwest region (Portland, Oregon/southern Washington). More recently, newer KP research centers have joined in multiregional and other multicenter pharmacoepidemiology studies. A hallmark of these studies is the ability to use computerized databases to identify patients exposed to pharmaceuticals of interest as well as appropriate unexposed comparison groups, and to measure and adjust for many confounding differences between these groups.

DESCRIPTION

Within KP, essentially all primary and specialty care and the vast majority of emergency and hospital care are delivered by providers belonging to a single medical group and working within a single care system for patients of a single health plan. All clinical information from each encounter is therefore captured accurately in clinical databases shared by providers, the health plan, and researchers. Each KP subscriber in every region receives a unique medical record number for all encounters with the program. This makes it straightforward to link patient records across databases (e.g., pharmacy records with hospitalizations, outpatient laboratory results, or claims received from non-KP providers) and across time.

The KP Membership

KP member populations are diverse, representative of the communities from which they are drawn, and quite stable. Although race/ethnicity is not routinely collected in any region, various sources, including member surveys and certain medical records, provide estimates that demonstrate rich racial and ethnic diversity in nearly every region. Unlike many for-profit, network model health plans, the proportions of KP members aged 65 years and above are close to population proportions, which is important when studying drugs used for chronic illnesses. In Northern California, KP members appear to be remarkably similar to the general population in terms of race/ethnicity in comparisons based on the residential addresses of members and census block group data from 1990 and 2000. KP members slightly under-represent those at the extremes of household income.

Health status of KP members is also thought to be similar to that of the general population. More than 90% of commercial subscribers along with their covered family members join KP through employer groups, with no prequalification screenings required, and the majority of KP members with Medicare (approximately 12%) “age in” to Medicare

coverage from commercial KP coverage. These membership features help to reduce potential differences in health status between members and non-members.

KP members tend to remain in KP for long periods, especially after the first one to two years of membership, making this population attractive for studies that require long-term follow-up (or follow-back in case-control studies). This is, in part, because KP physicians are not accessible except through KP health insurance; patients must remain within KP to maintain relationships with their personal physician. By contrast, patients in network model HMOs often have to switch health plans to remain with the same physician. Within KP Northern California, approximately 12% of all members depart within one year and 20% leave within two years. Thereafter, fewer than 5% leave per year, so that at 10-year follow-up more than 50% of the initial cohort remains enrolled and under observation. Retention is much higher for older patients and those with chronic illness.

Although dropout due to departures from the health plan may be modestly greater than that in epidemiologic cohorts of volunteer patients, it is also important to note that over 90% of plan dropout results from loss of access to KP membership due to changes in employment or employers’ decisions to discontinue KP coverage, rather than from active decisions by members to drop out of KP. This decreases the risk that dropout is related in some way to risks for outcomes (i.e., informative censoring).

KP Clinical and Administrative Databases

Administrative and clinical data sets are maintained in every region and used for clinical care, payment, and operational purposes. These include membership databases, demographic databases, hospitalizations, outside referrals and claims, outpatient visits, laboratory use and results, prescriptions, immunizations, service costs, cancer and diabetes disease registries, multiphasic health check-ups, member health surveys, acute coronary syndromes registries, neonatal minimum data sets, linked mortality databases, 2000 geocode databases, neurodegenerative disease registries, linked birth databases, adverse and allergic drug event reporting databases, immunization databases, genetics registries, dental administration and clinical tracking systems, perinatal databases, health risk appraisal databases, and operational data stores with vital signs data. Because of the size of the KP membership, these databases provide remarkably large and well-characterized study populations for addressing numerous pharmacoepidemiologic questions. Membership databases allow identification and follow-up of patient cohorts by age, sex, and area of residence, and immediate censoring of 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.

Pharmacy databases capture the vast majority of all prescription drug use in KP members, since well over 90% have pharmacy prescription coverage. In two regions (KP Northwest and KP Colorado), prescriptions can be identified at the time they are ordered. In all other regions, capture does not occur until the prescription is filled.

Uniform hospital discharge records are available in each region and have been used as a source of outcomes data for many years in KP studies. For endpoints not already studied and validated, chart review is often performed to confirm diagnoses.

Laboratory tests with CPT-4 procedure codes and results are valuable for assessing disease severity, physician laboratory monitoring practices, and dosage modification in the presence of laboratory abnormalities. They may also be useful for identifying certain endpoints (e.g., new liver function test abnormalities) in patients on specific medications. However, because tests are not performed routinely and regularly in clinical practice, data for specific tests will be missing for significant fractions of most populations.

Outpatient visit counts, by department and type of provider, are useful in studies of utilization patterns and costs of care associated with use of specific medications. Outpatient diagnoses are the most important data source for identifying patients with a disease and for measuring and adjusting for levels of comorbidity (case-mix). However, the validity of outpatient diagnostic data is not as well documented as for inpatient diagnoses. Several studies indicate that, when present, these diagnoses are highly indicative of the presence of the stated illness, but there is little information on the sensitivity of these databases. For this reason, outpatient diagnoses have been used relatively rarely as a source of outcomes.

In both Northern and Southern California, data for past and present members are linked to California death certificates using the following identifiers: social security number, name, date of birth, ethnicity, and place of residence. Linkage programs assign probabilistic weights to each purported match, allowing users to choose how conservative to be in accepting matches as valid. Researchers in KP Northwest also link member data with state vital statistics (e.g., birth and death) records for both Oregon and Washington. In each region, these data are valuable for studies of cause-specific and total mortality.

Additional Research Databases used by KP

Individual KP research centers have developed additional databases for research studies, including many condition-specific disease registries. Most are updated regu-

larly and can provide efficient approaches for studying questions related to the natural history of these conditions, the effectiveness or adverse effects of medications, and treatment patterns. For example, complete cancer incidence data for KP members are captured in registries maintained by the research departments in at least four regions. In both California regions, registries are linked to the California State Cancer Registry. In the Northwest and Colorado regions, SEER-compatible registries have been approved by the National Cancer Institute for research purposes. Data are collected in a standardized format no later than six months post-diagnosis. Diabetes registries are also available in at least these same four regions and have been in place for 10 years or longer. These registries have been shown to have extremely high sensitivity and positive predictive value for diabetes. Together, these four registries count more than 400 000 currently enrolled diabetic patients. The KP Northern California HIV/AIDS registry captures data on all members that meet diagnostic criteria for HIV infection. Verification and collection of additional information by medical record review is then performed for each potential case.

The KP geocoded membership database for Northern California links residential addresses for 2 658 488 members who were active and had mailing addresses in the primary 14-county catchment area of Northern California as of January 1, 2000 with US Census block group data on socioeconomic status and race/ethnicity. These data can be used as adjusters for socioeconomic status in comparisons of outcomes for users versus nonusers of drugs of interest. A cost-accounting database in KP Northern California also provides estimates of fully allocated costs by clinical department and by unit of service by integrating utilization databases with the program's general ledger. These data are very useful for comparing total utilization and costs of care between patient groups.

The multiphasic health check-up was a physical examination and extensive interview administered to more than 500 000 KP members at two Northern California medical centers between 1964 and 1984. Interview responses and physiological and laboratory results were computer stored and have provided a rich database on a cohort that included over 60% of adult members enrolled at the two centers. This database, often linked with subsequent outcomes provided by other data sources, has been the source of well over 200 publications over the past 35 years, and remains useful as a source of baseline information in retrospective cohort studies, particularly of older medications.

KP Northwest was the first region to implement an electronic medical record, EpicCare®, covering all outpatient care since 1997. EpicCare® describes the clinical care of more than 900 000 unique KP Northwest members through

December 2003. It also captures types of encounters not included in these databases (e.g., telephone consults), and includes more detail, such as provider orders for prescriptions or laboratory tests, regardless of whether patients decide to act on the order. This feature can provide insight in studying questions of the quality of care and safety in large populations. EpicCare® also captures full-text clinical notes, which can then be searched by visual chart abstraction, computerized search for text words, or computerized search using natural language processing algorithms to identify more complex patterns. Because EpicCare® is updated daily, incident disease can be identified rapidly for administration of surveys or telephone interviews in studying episodes of illness or natural history of disease.

STRENGTHS

By virtue of the size, diversity, representativeness, and relative stability of its membership and the increasing richness of its computerized clinical data, KP is an appealing site for conducting pharmacoepidemiologic studies. The KP membership or selected patient subgroups can often be thought of as cohorts with rich clinical information. The key computerized databases—membership, pharmacy utilization, laboratory results, and outpatient diagnoses—cover essentially the entire enrolled populations and have now been in place for at least 10 years. Thus, cohort studies with considerable follow-up (and case-control studies with similar lengths of follow-back) are feasible. (See Case Example 12.2.)

WEAKNESSES

Several weaknesses, including member dropout rates that are somewhat higher than those in studies of volunteers, are discussed above. Another important disadvantage is the absence of complete, standard information on race/ethnicity or other indicators of socioeconomic status for all members. Certain databases, including hospital discharge data and cancer and HIV/AIDS registries, routinely collect race/ethnicity. Data sets constructed by primary data collection in previous studies also contain this information. These data sources sometimes can provide cohorts with complete data and sufficient size to address certain research questions.

A continuing limitation of outpatient diagnostic databases is incomplete capture of all outpatient diagnoses, particularly for those not listed on specialty-specific encounter forms. This concern is reflected in the absence of studies using these databases as the primary source of outcomes. However, these outpatient databases remain extremely useful for initial construction of patient cohorts to study treatment/outcome associations and for case-mix adjustment.

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. As with most managed care formularies, KP formularies are somewhat restrictive, with one or two agents from a particular drug class being used almost exclusively. Newer agents may also be somewhat slower to achieve widespread use than in the fee-for-service environment. This hampers head-to-head comparisons of related drugs for effectiveness and toxicity.

By 2006, KP plans to implement a full electronic medical record, including physician order entry and clinical decision support systems, as well as scheduling and billing software for all inpatient and outpatient settings that will preserve the present capabilities described here, but will also bring greater uniformity to much of the data across KP's eight regions, making cohort identification and pooling of follow-up experience across regions more complete and efficient.

CASE EXAMPLE 12.2: TROGLITAZONE USE AND RISK OF HEPATIC FAILURE

Background

- The thiazolidinedione class of oral anti-diabetic medications represents a novel and important new approach to controlling blood glucose in diabetes.
- Shortly after introduction of the first thiazolidinedione, troglitazone, in 1997, spontaneous reports of acute hepatic failure in users of this agent began to appear, including death. As the number of reports increased, and as alternative drugs in the class without this effect emerged, the FDA and the drug's manufacturer agreed to withdraw the drug from the market in 1999.
- No controlled epidemiologic studies were available at the time of this decision to help quantify the absolute or relative increase in risk associated with troglitazone.
- Also, diabetes itself, particularly when poorly controlled, is known to increase risk for hepatic failure.

Question

- How does the risk of hepatic failure in troglitazone users compare with that of other diabetic patients?

Approach

- Researchers at the Kaiser Permanente Northern California Division of Research and collaborators from

(Continued)

four other members of the HMO Research Network (HealthPartners, Fallon Community Health Plan, Harvard Pilgrim Health Care, and Lovelace Foundation) used a retrospective cohort design to identify a cohort of more than 170 000 adult diabetic patients.

- The cohort included over 9600 troglitazone users with over three years of drug exposure.
- Hospital discharge diagnoses and procedures potentially indicative of acute liver injury were identified, and the full-text medical records of over 1200 possible incident events were reviewed.
- Medical records of 109 cases were sent to a panel of hepatology specialists for blinded adjudication.

Results

- The panel ultimately identified 35 cases of acute hepatic injury or failure not clearly attributable to a known cause other than use of diabetes medications.
- Risk in troglitazone users was not found to differ from that of other diabetic patients.
- However, the entire diabetes cohort was at increased risk compared with the general population.
- Acute liver failure or injury not clearly attributable to other known causes occurred on the order of 1 per 10 000 person-years among diabetic patients treated with oral hypoglycemic drugs or insulin.
- This study strongly suggested that any troglitazone-related increase in risk was much smaller than the 20–25-fold increase suggested by spontaneous reports data.

Strengths

- This cohort study utilized a well-defined source population.
- This study design permitted calculation of age- and sex- standardized incidence of liver disease.
- For the subset of study patients who had at least one year of health plan membership before the first dispensing of a hypoglycemic agent during the study, it was possible to calculate a comorbidity score using the pharmacy dispensing records for these patients to identify chronic disease. This made possible adjustment for comorbidity in the analysis.
- Use of a structured medical record review with explicit criteria for many of the exclusions and use of a masked adjudication process helped to eliminate bias.

Limitation

- Although this study had the largest reported cohort of troglitazone-exposed persons, cohorts with even more troglitazone users are needed to more precisely estimate the incidence of these rare events.

Summary Points

- The incidence of acute liver failure or injury in troglitazone users was similar to that among patients exposed to different groups of hypoglycemic agents.
- Very large cohort studies are required to precisely estimate the incidence and risk of rare diseases.

HMO RESEARCH NETWORK

The HMO Research Network (<http://www.hmoresearchnetwork.org>) is a consortium of 15 health plans. It advances population-based health and health care research in the public domain by using health plans' defined populations, their clinical systems, and their data resources to address important medical care and public health questions. Each of the health plans is home to a research unit that develops and implements its own research portfolio. In addition, these research groups work together through various formal and informal collaborations, including the HMO Research Network Center for Education and Research on Therapeutics (CERT) funded by the Agency for Healthcare Research and Quality (AHRQ), the Cancer Research Network funded by the National Cancer Institute, an Integrated Delivery Systems Research Network funded by AHRQ, a Collaborative Clinical Studies Network funded by the National Institutes of Health (NIH) as an NIH Roadmap contract, and the Vaccine Safety Datalink of the National Immunization Program.

We will discuss here the data sources that HMO Research Network members utilize through their participation in the CERT, created in response to a congressional mandate in 1999. The mission of the CERTs includes research and education to advance the optimal use of drugs, medical devices, and biological products (<http://www.certs.hhs.gov>).

DESCRIPTION

Member Health Plans

The 15 members in the HMO Research Network are Geisinger Health System in Pennsylvania; Group Health

Cooperative (discussed earlier in this chapter) in Washington and northern Idaho; Harvard Pilgrim Health Care in eastern Massachusetts; HealthPartners Research Foundation in Minnesota; Henry Ford Health System—Health Alliance Plan in Michigan; Kaiser Permanente (also discussed earlier) Colorado; Kaiser Permanente Georgia; Kaiser Permanente Hawaii; Kaiser Permanente Northern California; Kaiser Permanente Northwest in Oregon; Kaiser Permanente Southern California; Lovelace Health System in New Mexico; Marshfield Clinic Research Foundation in Wisconsin; Meyers Primary Care Institute/Fallon Healthcare in central Massachusetts; and Scott and White Memorial Hospital in Texas. As of 2006, nine of these members of the HMO Research Network participate in CERT. They serve geographically and ethnically diverse populations with a broad age range and relatively low turnover rate; together, the health plans have more than 7 million members with longitudinal data.

Data Elements in Automated Databases

Demographic Data and Membership Status

Date of birth and gender are routinely available. Race and socioeconomic status are imputed using census data by geocoding the addresses of health plan members and linking to 2000 US Census data. Each health plan maintains detailed information on dates of enrollment, termination of membership, and changes in benefit plans for each health plan member for billing purposes. This information is usually used to qualify and identify subjects with incident drug use in inception cohorts. Annual membership turnover rates are between 10% and 15% for general health plan membership in most plans. Membership retention rate is much higher among patients with chronic diseases.

Drug Exposure

Approximately 90% of members have a pharmacy benefit that provides a strong financial incentive for them to receive their drugs through a mechanism that results in a claim for reimbursement. Each drug dispensing record contains a unique National Drug Code (NDC) that identifies the active ingredient(s), dose, formulation, and route of administration. Amount dispensed, days of supply, and prescribing physicians are also included in the dispensing records. Based on the NDC dictionary published by the FDA and commercial data sources, a comprehensive drug dictionary has been built and is updated regularly for research purposes. Drugs 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 codes, for the particular office visit during which the drug was administered. However, ascertainment of drug exposure using these automated records may not provide a complete picture. Information on drugs that are used during hospitalizations is usually not available. Some benefit plans have an annual drug expenditure limit beyond which medications have to be paid out-of-pocket by the member. The health plans typically have no record of dispensing for drugs after the limit has been reached, and do not have information on the use of over-the-counter medications.

Diagnoses and Procedures/Special Examination

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

Medical Records

In addition to claims data, medical records for ambulatory visits are currently available electronically in six health plans. Most of the remainder are adopting electronic medical records. These automated records allow efficient access to vital signs, laboratory test results, prescribing data (versus dispensing data from claims), and full-text clinician notes. For all health plans, full-text medical records in paper form are either available within the health plan or requested from hospitals and other medical care providers.

Linkage to External Registries and Data Sources

Health plan data can be linked with external data for research purposes. For example, linkage to cancer registries and the National Death Index has been performed to ascertain cancer and mortality outcomes

STRENGTHS

Large cohorts from diverse populations and delivery systems can be identified to evaluate rare events incidence and to study these events among sufficiently large numbers of patients with certain comorbidities, such as hypertension, diabetes, and congestive heart failure. The diverse ethnic composition of the health plans allows for studies that are more likely to be generalizable in the US than may be the case in single health plans. (See Case Example 12.3.)

WEAKNESSES

The general limitations of health plan-based data sources apply to the HMO Research Network. The most important of these are absence of population groups that are uninsured, underrepresentation in some HMOs of the elderly, turnover of the population, carve-outs of some services, caps on certain services, some constraints on formularies, and lack of information on potential confounders that are not captured in automated data or written medical records. While the data are rich in elements related to health care, those for race, ethnicity, and lifestyle factors such as smoking and alcohol consumption are not yet readily available.

CASE EXAMPLE 12.3: MEDICATIONS WITH BLACK BOX WARNINGS

Background

- Black Box Warnings are special sections in package inserts to communicate specific serious safety issues about a drug. Its lack of effectiveness has been demonstrated for cisapride, but there have been few studies on Black Box Warnings for other drugs.

Question

- What is the prevalence of use of drugs with Black Box Warnings in the general population and what is the level of compliance with the warnings?

Approach

- A retrospective study based on automated data from ten HMOs from January 1, 1999 through June 30, 2001 was conducted to evaluate the use of 216 drugs with Black Box Warnings.

Results

- During the 2.5-year study period, more than 40% of almost one million health plan members from geographically diverse regions had dispensing records of at least one drug with a Black Box Warning that might apply to them.
- Almost half of the apparent non-compliance to warnings were lack of baseline laboratory monitoring before initiation of the drug.
- Rates of apparent non-compliance with inappropriate concomitant medications were low.

Strength

- Efficient use of automated information from large population.

Limitations

- Investigators could only evaluate dispensing records and had limited information on the intended indication for the prescriptions.
- Investigators could not ascertain whether a prescriber was not aware of the Black Box Warning of the drug or the prescriber was aware of the warning but chose not to follow the suggestions.
- It was essentially a cross-sectional study. Longitudinal data before and after the announcement of Black Box Warnings are needed to evaluate the effectiveness of the warnings.

Summary Points

- We need more information on how the “high risk” drugs, empirically defined as drugs with Black Box Warnings, are used in the population.
- Rates of compliance with Black Box Warnings varied for different drugs.
- More empirical information is needed to evaluate how Black Box Warnings convey drug safety information and how it may change prescription behavior.

UNITEDHEALTH GROUP

UnitedHealth Group is a diverse company providing health and well-being services to more than 50 million members throughout the US. The company established the Center for Health Care Policy and Evaluation in 1989 as a private sector research institute with an independent research agenda. The Center has fostered objective public health research for UnitedHealth Group-affiliated health plans and specialty companies (such as Uniprise, which provides benefits programs to leading businesses in the US, and AmeriChoice, which provides public sector health care programs), as well as federal and private sector clients.

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, which provides UnitedHealth Group researchers and their collaborators with unique research opportunities. A subset of the electronic claims data, the research databases, have been used to conduct various customized studies, including pharmacoepidemiologic

research, analyses of health care quality, performance and outcomes, and economic evaluations, often in collaboration with academicians and/or government agencies.

DESCRIPTION

Overview of UnitedHealth Group

UnitedHealth Group (www.unitedhealthgroup.com) is a diversified health and well-being company serving consumers, managers, and health care professionals. Founded in 1974, the company serves more than 50 million persons through a continuum of health care and specialty services, which include point of service arrangements, preferred provider organizations, managed indemnity programs, Medicaid and Medicare managed care programs, and senior and retiree insurance programs. Other services include managed mental health and substance abuse services, care coordination, specialized provider networks, third-party administration services, employee assistance services, evidence-based medicine services, and information systems.

UnitedHealth Group-affiliated health plans presently reach over 16 million people across the US, and are located in all geographic regions, with urban and rural representation. Plan members 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. Over 50% of all hospitals in the US are part of the network.

Research Databases

The research databases that typically have been used for public health research consist of current and historical medical and pharmacy administrative claims data submitted by 11 UnitedHealth Group-affiliated geographically diverse health plans, situated in the northeastern, southeastern, midwestern, and western regions of the US. The administrative claims research databases are large longitudinal databases, containing more than 10 years of data from 1993 to the present. In 2002, there were approximately 3.8 million members and 2.8 million member-years, representing commercial, Medicaid, and Medicare populations. For the purpose of pharmacoepidemiologic evaluations for post-marketing drug surveillance, analyses typically are restricted to those members having a drug benefit. More than 90% of commercial members and most Medicaid members in the research databases have a prescription drug benefit. Since Medicare drug benefits vary depending on the plan, pharmacy files may not capture all prescribed drugs if Medicare

beneficiaries reach their drug benefit limits. The number of members in the research databases varies annually but is expected to continue to increase. The research databases are used to link files longitudinally at the individual level and are organized from the following components:

- *Membership data.* A member enrollment file stores demographic information on all health plan members, including dependents. Data elements include date of birth, gender, place and type of employment, and benefit package, as well as linkage to dates of enrollment and disenrollment. A unique identifier is assigned to each member at the time of enrollment and is retained if a member disenrolls and later re-enrolls. Precautions are taken to safeguard the confidentiality of individually identifiable information and Protected Health Information as required by state and federal regulations.
- *Medical claims.* A claim form must be submitted by a health care professional to receive payment for any covered service. Medical claims are collected from all health care sites (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 and other treatment. Claims are submitted electronically or by mail.
- *Pharmacy claims.* Claims for covered pharmacy services typically are submitted electronically by the pharmacy at the time a prescription is filled. The claims history is a profile of all prescription drugs covered by the health plan and filled by the member. Each claim specifies the pharmacy code, drug name, date dispensed, dosage of medication dispensed, duration of the prescription in days, and quantity dispensed.
- *Health professional data.* 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.

Research capabilities include:

- *Performing record and file linkages.* Enrollment, medical claims, pharmacy claims, and physician claims can be integrated by linking members' discrete records, inpatient claims, outpatient claims, and pharmacy claims to member and health professional data. These linkages allow for analyses of episodes of care and investigation of procedures and treatments regardless of care location.

- *Constructing longitudinal histories.* Information on diagnosis, treatments, and the occurrence of adverse clinical events, as coded on claims, can be tracked across time. Adherence to recommended patient laboratory and other testing and persistence of use for specific medications can be evaluated. To facilitate these processes, more complete longitudinal histories are constructed by tracking members who have had multiple enrollment periods and identification numbers within and across plans. Similarly, programs have been written to combine data from physicians with multiple identification numbers.
- *Identifying denominators to calculate rates.* The research databases can be used to calculate population-based rates, and to adjust resource use rates for the effects of partial-year enrollment. Through the member enrollment file, all individuals eligible to receive medical services or outpatient pharmacy services are identified. These populations can be defined by age, gender, benefit status, period and enrollment duration, or geography. Through the medical and pharmacy claims, subgroups of the membership can be identified for calculating the prevalence and incidence of specific diseases and conditions or use of particular treatments.
- *Identifying treatment at a particular point in time.* The ability to identify and track treatment is a critical function in pharmacoepidemiologic research. For instance, specific treatments can be identified using pharmacy and procedure codes.
- *Identifying cases and controls for study.* Programs have been developed and tested to identify and select cases and controls for study based on eligibility criteria such as insurance benefit status, age, current and/or continuous enrollment during a specific period of time, disease diagnosis, and covered medical procedures or drug therapies.
- *Identifying the treating physician.* For many studies, it is essential to attribute members' health care to a particular physician or other health professional. For example, researchers may want to locate a medical record for the collection of detailed clinical information not captured in the claims data. Because members receive care from multiple physicians, logic algorithms have been developed to identify the physician who provided the majority of care or key treatments for a particular medical condition during the study period of interest.
- *Calculating person-time at risk and time of event occurrence.* The databases contain the data elements necessary to calculate person-time at risk, i.e., the date on which the prescription was filled, the amount dispensed, the duration of the prescription (days' supply), and the period and duration of enrollment for each

member. The drug strength, amount dispensed, and days' supply fields can be used to estimate the total dose per prescription, the cumulative dose, or the time-at-risk above a recommended dose. Software has also been developed to calculate the number of days that members have been enrolled in the plan.

- *Obtaining medical record abstractions for validation purposes.* The current process to abstract medical records has been developed in collaboration with UnitedHealth Group-affiliated health plans, and has been successfully used in a number of studies. Although time consuming, the process is designed to ensure efficiency and data integrity, protect data confidentiality, maintain the health plans' relationship with its providers, and minimize provider burden.
- *Evaluating the impact of risk communication efforts.* Successful government management of drug safety risks requires effective communication with health care providers and the public. The evaluation of the effectiveness of risk communication efforts requires data on real-world medical practice. The research databases provide nationally representative information on actual clinical practice for these evaluations, for example whether testing recommended by the FDA labeling is completed prior to exposure to a new drug.

STRENGTHS

The research databases provide an efficient and unobtrusive method to identify and study exposures to prescription drugs and biologics. Further, drug or biologic utilization and potential adverse events identified within the plans reflect usual practice within the general medical community. Whereas clinical trials for new drugs are typically conducted in university or other unique settings to determine efficacy, postmarketing surveillance enables researchers to determine the effectiveness when usage has diffused to the broader medical community.

The varied types of plan locations and differences in characteristics of populations, in combination with the large size of the databases, provide unique advantages to conduct pharmacoepidemiologic research studies. (See Case Example 12.4.) The diverse demographic characteristics of members increase the utility of these databases for such research. Specifically, populations of children (almost 1 million in 2000), pregnant women (approximately 32 000 deliveries in 2000), and the elderly (over 210 000 were 65 years of age or older in 2000) in the research databases are sufficient for most analyses. The use of these standardized databases provides numerators and denominators for exposures to drugs and allows the estimation of incidence and

prevalence. Further, rare exposures and rare outcomes can be detected given the size of the databases.

The databases also provide the ability to link various types of files longitudinally for individual members regardless of service site. Thus, adverse events and outcomes may be analyzed by considering temporality in relation to drug exposure through pharmacy claims and linking such varied health services as hospitalization, emergency department use, physician visits, and any other site of health care service. Further, with respect to pharmacy claims, the prescribing physicians can be determined as well as their specialty and location. Claims submissions are generally complete, since claims must be submitted by health professionals for payment in most of the health plans.

Administrative claims data are useful for quality assessment and as a screening tool to identify quality problems. Similarly, the databases can be used to identify cases and controls or cohorts for study and additional information then can be obtained from medical records. This supplemental information can be crucial in pharmacoepidemiologic research studies. For example, through the abstraction of medical records, one can confirm a diagnosis and obtain additional information on risk factors and outcomes.

WEAKNESSES

There are certain structural constraints that limit access to obtaining all possible prescription drug claims. Like many other resources for pharmacy management services, data on the use of inpatient drugs are not available. In addition, given the pharmacy benefit structure, if the cost of a prescription drug is lower than the copayment amount, the prescription may not be included in the database since no prescription claim may be submitted. Overall, if a drug is not covered on the preferred drug list, exposures to that specific drug may be limited since the copayment is higher. Although drug exposure is sizable, these omissions have implications both for sample size and controlling for confounding by the omitted drugs.

A limitation of claims data with respect to characterizing exposure to drugs is a lack of information on patient adherence with the therapeutic regimen. However, several 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.

With respect to the completeness of the databases, some plans that have different financial incentives from the typical discounted fee-for-service mechanism may not have complete data. 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 databases. This disadvantage may be addressed by excluding this small number of plans from data extraction for research studies.

Another limitation is the length of time required to obtain all claims for a given time frame. The claims lag is short for pharmacy claims (1 month) but longer for physician and facility claims (7–8 months). The claims lag may be variable across study years and should be taken into account in study design.

CASE EXAMPLE 12.4: EVALUATION OF THE ASSOCIATION BETWEEN TROGLITAZONE AND ACUTE LIVER FAILURE

Background

- Cases of acute liver failure (ALF) in patients taking troglitazone for type II diabetes were reported to the FDA soon after initial marketing in 1997. Although the FDA convened an advisory committee in March 1999 to review data on 43 US cases of ALF, the drug remained on the market for another year (by which time 94 cases of ALF had been reported).

Question

- Although the number of spontaneous reports of ALF associated with troglitazone treatment was substantial, what were the incidence rates of idiopathic ALF and hospitalized acute liver injury in a large, inception cohort of troglitazone users?

Approach

- A study was performed at the Center for Health Care Policy and Evaluation in collaboration with the Center for Drug Evaluation and Research of the FDA to answer the above question.
- Administrative claims data from 12 UnitedHealth Group health plans were used to identify members with at least one troglitazone prescription between April 1, 1997 and December 31, 1998, and a minimum of 90 days of continuous enrollment before their index (first) troglitazone prescription.
- Person-years of exposure to troglitazone were calculated for each patient based on the cumulative days' supply of all prescriptions filled during the observation time period.

(Continued)

- Hospital claims data were searched to identify hospitalizations indicating a discharge diagnosis of liver disease or procedures suggesting possible liver disease occurring after the date of the index prescription. Claims data for other conditions that might explain the disorder were also examined.
- For patients whose administrative claims data were suggestive of ALF or inconclusive, hospital medical records were reviewed. For a diagnosis of ALF, hepatic encephalopathy, liver transplantation, or death in the setting of acute, severe liver injury was required.

Results

- 7568 patients with 4020 person-years of exposure to troglitazone were identified for the inception cohort.
- Nineteen patients had a liver-related hospitalization after their index prescription, of whom ten were excluded on the basis of the review of their claims data and a further four were excluded after reviewing their medical records.
- The estimated incidence rates of acute idiopathic liver injury per million person-years, with 95% confidence intervals, were 1244 (404–2900) for hospitalization ($n = 5$), 995 (271–2546) for hospitalized jaundice ($n = 4$), and 240 (6–1385) for ALF ($n = 1$).

Strengths

- Ability to identify a large population-based cohort of patients dispensed the drug of interest and to determine a rare potential adverse effect.
- Availability of clinical information about the potential adverse effect to confirm the electronic data.
- Results from a population-based cohort study allow a risk assessment that is not possible from spontaneous reporting information.

Limitations

- Analyses based only on electronic data may include recording errors (a claims error was found for one patient in the study).
- Not all medical records may be accessible (the record for one patient in the study).

Summary Points

- Administrative health care utilization data can be used to rapidly identify a cohort of patients that have been dispensed a drug of interest and to provide health care outcome information to evaluate drug safety signals.

- Access to medical records often may be required to confirm claims information.
- The results allow a population-based evaluation of drug safety that is not available from spontaneous reports submitted to pharmacovigilance agencies.

MEDICAID DATABASES

Medicaid is currently the largest US government-funded program that pays for both outpatient prescription drugs and medical care. Medicaid data have been used for pharmacoepidemiologic research since the early 1980s and continue to be used actively today.

DESCRIPTION

Description of the Medicaid Program

Medicaid is funded jointly by the federal government and individual state governments and administered by states with federal oversight. Medicaid provides benefits to US citizens and lawfully admitted immigrants only if they belong to one of three general groups: (i) low-income pregnant women and families with children, (ii) persons with chronic disabilities, and (iii) low-income seniors, including those receiving Medicare benefits. Medicaid programs function as a payer rather than as a direct provider of health care services.

Characteristics of Medicaid Recipients

In 2002, 51 million persons, or 16% of the US population, received health care services through Medicaid. Children, females, and nonwhites are overrepresented.

Sources of Medicaid Data for Research

The Centers for Medicare and Medicaid Services (CMS) is a major source of Medicaid data for researchers. CMS receives data from individual state Medicaid programs, and performs extensive editing, range checks, and comparisons with previous data from that state when preparing data files. There is currently a lag of approximately four years between the end of a calendar year and when data from that year become available from CMS. Since 1997, the University of Minnesota School of Public Health's Research Data Assistance Center (ResDAC) has, through a contract

with CMS, provided free assistance to academic, government, and nonprofit researchers interested in using Medicaid and Medicare data for research. The data are obtained from CMS.

Pharmacoepidemiologic research has also been conducted using data obtained directly from individual states, including California, New Jersey, New York, and Tennessee.

Data Structure

CMS provides Medicaid data in five distinct types of data files: Personal Summary, Inpatient, Prescription Drug, Long Term Care, and Other Therapy. There is one of each file type for each state for each calendar year.

The Personal Summary file contains one record per individual enrolled in that state's Medicaid program for at least one day during the relevant year. This file includes demographic data, identifies which months the person was enrolled in Medicaid, and which months (if any) the person participated in a managed care plan.

The Inpatient file contains information on hospitalizations. Available information includes a code identifying the hospital, admission date, discharge date, discharge status, up to nine diagnoses (coded in the International Classification of Diseases, 9th edition, Clinical Modification; ICD-9-CM), up to six procedures (coded in Current Procedural Terminology CPT-4, ICD-9-CM, or other coding systems), and payment information. The Inpatient file does not contain information on drugs received in the hospital. Therefore, Medicaid data are not useful for studying inpatient drug exposures.

The Prescription Drug file contains one record for each reimbursed outpatient or nursing home prescription. Drugs are coded in a non-hierarchical system known as the National Drug Code (NDC). In addition to the NDC, records in the Prescription Drug file also include the date the drug was dispensed, the identification number of the prescriber (although this field is often missing), the quantity dispensed (e.g., number of tablets), whether the prescription was new or a refill, cost information, and the intended duration as estimated by the pharmacy ("days' supply").

The Long Term Care file contains encounter records for long-term care services provided by skilled nursing facilities, intermediate care facilities, and independent psychiatric facilities. Fields include facility type, dates of service, diagnosis, and discharge status.

The Other Therapy file contains encounter records for all non-institutional Medicaid services, including physician services, laboratory, radiology, and clinic services. Capitation payments for persons in capitated managed care plans are also included. For laboratory and radiology encounter

records, the type of test, but not its results, is reported. Date and type of service and, if applicable, diagnosis or procedure code are reported.

Other types of data have been linked to Medicaid data to enhance its utility. For example, Medicare data have been linked to increase the proportion of medical care encounters identified among individuals eligible for both Medicare and Medicaid. Such a link can be very important, since Medicaid encounter records can fail to document a considerable proportion of care provided to individuals who are simultaneously eligible for both Medicare and Medicaid. Medicaid data have also been linked to mortality data such as the Social Security Administration Death Master File, National Death Index, and state vital statistics registries. Finally, cancer registries have been linked to Medicaid data to study possible carcinogenic effects of medications. The accuracy and completeness of such linkages cannot be assumed, but rather need to be evaluated by researchers relying on them.

STRENGTHS

An important strength of Medicaid databases is their large size, which permits the study of infrequently used drugs and rare outcomes. More than 10 states have over a million Medicaid recipients each.

Another strength of Medicaid data is that the outpatient prescription encounter records accurately record the date, NDC, and quantity dispensed by the pharmacy for each Medicaid beneficiary. Because this information determines the payment provided by the Medicaid program to the pharmacy, it is subject to regulatory audit and has been shown to be highly accurate. Because patients have a financial incentive to use Medicaid to purchase their drugs instead of paying for them out-of-pocket, it seems reasonable to expect that pharmacy encounter records capture the vast majority of prescription drug use, particularly in the low-income populations served by Medicaid.

Like prescription records, encounter records that code for clinical procedures determine the amount paid to the health care provider. Therefore, procedure records are audited to detect fraud, and thus would be expected to be highly accurate with regard to performance of that procedure.

Another potential strength of Medicaid is overrepresentation of certain populations. Medicaid has substantially greater proportions of pregnant women, young children, and African Americans than other data sets. Approximately 11% of Medicaid beneficiaries are 65 years of age or older. Because populations overrepresented in Medicaid are often underrepresented in randomized trials, the opportunity to study them in Medicaid is particularly valuable.

WEAKNESSES

Generalizability

Medicaid recipients are not representative of the general population with respect to age, race, income, and disability status. Therefore, some results may not generalize well to the broader population. This is particularly true of descriptive studies of health care utilization. However, the generalizability of etiologic studies is compromised only for biologic relationships that vary based on factors that differ between Medicaid and non-Medicaid populations. Thus, the results of etiologic studies have generally been consistent between Medicaid and non-Medicaid populations.

Diagnostic Terminology

Diagnoses in Medicaid are coded using the ICD-9-CM coding scheme, which can be problematic for researchers for several reasons. First, there are often many ICD-9-CM codes compatible with a single clinical condition. Therefore, it is usually necessary to include many different ICD-9-CM codes to identify an outcome of interest.

Second, there is no incentive for providers to use the most specific code available. Thus, one must be careful not to over-interpret diagnostic distinctions within claims data. Generally, researchers using claims data should be “lumpers” rather than “splitters.”

Third, ICD-9-CM codes do not always precisely fit the clinical condition of interest. Therefore, several sets of codes are often needed to define the same disease.

Fourth, the ICD-9-CM coding scheme is often too non-specific for research purposes. When coded encounter diagnoses are not specific enough, researchers must use primary medical records to distinguish among clinical conditions.

Finally, outcomes that do not reliably result in encounters with health care providers may be under-ascertained. This can introduce bias into the findings of analytic studies if the likelihood that a given clinical condition leads to medical care is related to use of a particular medication.

Limitations in Prescription Coverage

Only drugs covered by Medicaid can be studied using Medicaid encounter data. Numerous drug categories are generally not covered by Medicaid, such as agents for fertility, weight loss, hair growth, cosmetic effect, and smoking cessation. Therefore, these agents cannot be studied using Medicaid data. Many states also require prior approval before reimbursing for certain drugs, such as human growth hormone, non-sedating antihistamines, and more expensive NSAIDs. The coverage of injectable drugs and adult vaccines also

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. Finally, states vary in their coverage of non-prescription drugs.

Data Validity

Probably the most important concern with using Medicaid data for research is its validity. A study funded by the FDA and performed by the Research Triangle Institute (RTI) in the early 1980s compared Medicaid encounter data from Michigan and Minnesota to its primary sources, i.e., clinical records in hospitals, physician offices, pharmacies, etc. The results of this study suggested that the demographic and drug data appeared to be of extremely high quality and the date of dispensing of prescriptions agreed in 97% of prescriptions, but medical diagnoses need to be validated by medical records review. The validity of diagnostic data must be considered in the context of each individual study outcome.

Several levels of validity of diagnosis data need to be considered when using Medicaid encounter data. The first is whether the encounter diagnosis accurately reflects the clinical diagnosis listed on the medical record. The second level is whether the clinical diagnosis made and recorded by the physician is correct.

For 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, to characterize the severity of the disease, and to obtain information on potential confounding variables not found in the encounter data. Potential exceptions are studies of outcomes for which encounter diagnoses have previously been found to be sufficiently valid, and studies using a procedure or a prescription for a drug as the outcome of interest. Examining the validity of diagnosis codes requires review of clinical records, which must frequently be done without subject contact, since contacting Medicaid beneficiaries who experienced specific outcomes years prior may be impossible or impracticable. Under the Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA), researchers with necessary documentation can legally request the hospital records of specific patients even without patient contact (see Chapter 19). Necessary documentation includes waivers of informed consent and of HIPAA authorization granted by an institutional review board, and a data use agreement with the agency providing the encounter data (e.g., CMS or the individual Medicaid agency). In most circumstances, hospitals would be required to record (in HIPAA parlance, “account for”) such disclosures of protected health information, and report those disclosures to any of their patients

who were subjects in the study and who request this information. Presently, there is little experience upon which to gauge the willingness of hospitals to provide researchers with access to medical records data using this mechanism. Unfortunately, if investigators are unable to obtain clinical records because of regulatory impediments, the utility of Medicaid encounter data to improve Public health will be compromised.

Identifying Enrolled Person-time

Studies using Medicaid data can validly include only person-time during which subjects would have had health care services reimbursed by Medicaid if they had occurred. Thus, the need to identify *enrolled person-time* is crucial, as investigators need to distinguish periods of health from periods of ineligibility. In studies that follow each subject for the expected duration of a given prescription, only a small proportion of subjects would be expected to become ineligible during the short follow-up period following the prescription. Thus, this issue may be relatively unimportant in this context. In contrast, this issue may be more important in case-control studies, in which it is necessary to identify a representative sample of person-time in the source population that gave rise to the cases, and in cohort studies with unexposed comparison groups.

One way to identify enrolled person-time is to use the information from the Personal Summary (i.e., enrollment) file, which lists the Medicaid enrollment dates for each subject enrolled in the program for each month during that year. There are two potential problems with relying on this information, however. The first is that, 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, states have been required to provide CMS with encounter data for individuals enrolled in capitated plans. However, despite this requirement, encounter data for those enrolled in capitated plans appear 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. The second potential problem is that anecdotal experience has suggested that enrollment information from some states may be inaccurate.

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., six months) both before and after the person-time under study. Naturally, this approach will miss fatal outcomes. It

also is unable to differentiate periods of ineligibility from periods of health.

In summary, very large studies can be performed with Medicaid databases in a relatively quick and inexpensive way. (See Case Example 12.5.) These databases permit studies of both inpatient and outpatient diseases, and sometimes permit calculation of incidence rates. A major concern in using this type of database is the validity of the diagnosis data. Thus, the ability to obtain clinical records to validate encounter diagnoses is crucial. Provided that appropriate steps are taken, health care providers are permitted under the HIPAA to provide investigators with access to clinical records even without patient contact. However, current experience does not allow us to gauge the willingness of providers to do so. This is a problem that must be resolved if these resources are to continue to demonstrate maximum utility.

CASE EXAMPLE 12.5: **USE OF MEDICAID DATA TO STUDY** **DRUG-INDUCED SUDDEN CARDIAC DEATH** **AND VENTRICULAR ARRHYTHMIA**

Background

- There is concern that antipsychotic drugs that prolong the electrocardiographic QT interval at standard doses (e.g., thioridazine) may be associated with a higher rate of sudden cardiac death and ventricular arrhythmia than antipsychotic drugs that do not prolong the QT interval at standard doses (e.g., haloperidol).

Questions

- Is thioridazine associated with a higher rate of sudden cardiac death and ventricular arrhythmia than haloperidol, either overall or at high dose?
- Does the risk increase with dose?

Approach

- Cohort study in a large database of Medicaid data.
- The rate of sudden cardiac death and ventricular arrhythmia was measured in patients receiving thioridazine and haloperidol.
- Confounding factors present in Medicaid data were measured and controlled for.
- Subgroup and dose-response analyses were performed.

(Continued)

Results

- There was no overall difference in the rate of sudden cardiac death and ventricular arrhythmia. In particular, the rate ratio (95% confidence interval) for thioridazine versus haloperidol was 0.9 (0.7–1.2).
- However, at doses of 600 mg per day or its equivalent, the rate ratio (95% confidence interval) for thioridazine was 2.6 (1.0–6.6). A dose–response relationship was evident for thioridazine but not for haloperidol.

Strengths

- The study compared two drugs used for the same or similar indications, reducing the potential for confounding by indication.
- The study was large, allowing meaningful comparisons of very rare events.

Limitations

- There may be both under- and over-ascertainment of sudden cardiac death and ventricular in administrative claims data.
- This study does not demonstrate the absolute safety of haloperidol, but rather the relative cardiac safety of the two drugs.
- The potential for unmeasured confounding cannot be excluded.

Summary Points

- Large sample sizes are needed to study rare outcomes.
- Confounding by indication is more likely to be a concern when comparing treated versus untreated subjects than when comparing groups receiving treatments for the same or similar indications.

HEALTH SERVICES DATABASES IN SASKATCHEWAN

Saskatchewan is one of ten provinces and three territories in Canada. It has a relatively stable population of about one million people, or about 3.2% of the national population.

Saskatchewan has a publicly funded health system. Within this system, Saskatchewan Health, a provincial government department, and 13 health authorities provide health services to the citizens of the province. In almost all of the provincially funded programs, residents of the

province enjoy universal health insurance. There is no eligibility distinction based on socioeconomic status.

As a by-product of these universal health care programs, the province maintains a large amount of health care information in computerized databases over many years. These databases are a recognized resource for pharmacoepidemiologic, drug utilization review, health economics, and other health services research.

DESCRIPTION

The major databases include the registry of the eligible population, prescription drug data, hospital services data, physician services data, the cancer registry, and vital statistics data. The databases include all segments of the population, including children, women of childbearing age, and the elderly.

Eligible Population

Saskatchewan residents are entitled to receive benefits through the health care system once they have established residence and have registered with Saskatchewan Health. A Health Services Number (HSN), assigned at registration, is a lifetime number that uniquely identifies each beneficiary. The HSN is captured in records of health service utilization and enables linkage of the computer databases.

Saskatchewan Health maintains an accurate, comprehensive, and current population registry that includes all residents eligible for health coverage. As of June 30, 2003, approximately one million people were eligible for health benefits. Excluded from eligibility, and therefore from the population registry, are people whose health care is funded fully by the federal government. (This category accounts for less than 1% of the total population.)

The population registry captures demographic and coverage data on every member of the eligible population.

Prescription Drug Data

All Saskatchewan residents are eligible for benefits under the Prescription Drug Plan, with the exception of approximately 9% of the population (primarily Registered Indians) who have their prescription costs paid for by another government agency.

Drugs covered by the Drug Plan are listed in the Saskatchewan Formulary which is published annually and updated quarterly. (As of July 2003, the Formulary included over 3500 drug products.) Data are captured on all prescriptions for drugs listed in the Formulary and dispensed to eligible beneficiaries. Data collected on each prescription

include patient, drug, prescriber, pharmacy, and cost information. The database contains information from September 1975 to June 1987 and from January 1989 to date. (Drug data are incomplete from July 1987 to December 1988.)

The drug database does not include information on most nonformulary prescription drug use, most over-the-counter (OTC) drug use, use of professional samples, or in-hospital medications. It also does not include information about the dosage regimen prescribed, the reason the drug was prescribed, days' supply or patient compliance.

Hospital Services Data

Under the hospital care insurance program, hospitals provide medically necessary services without charge to beneficiaries. All members of the covered population are eligible to receive benefits. Data are collected on every hospital separation (defined as discharge, transfer, or death of an inpatient) and day surgery case. These data are accessible electronically from 1970 to the present.

Data are collected from all acute care hospitals in the province. As well, the database includes information on out-of-province hospital separations involving a member of the covered population. Standard diagnostic and procedure classification systems are used. In data collected up to March 31, 2001, diagnoses were recorded using four-digit codes based on the ICD-9, and procedures using four-digit codes based on the Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures (CCP). Effective from April 2001, diagnoses and procedures are recorded using the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Canada (ICD-10-CA) and the Canadian Classification of Health Interventions (CCI). Diagnostic coding for the majority of hospital discharges is undertaken at the hospital level, usually by health records administrators.

Physician Services Data

In Saskatchewan, most physician services are an insured benefit. All members of the covered population are eligible to receive benefits. Medical, surgical, obstetric, anesthesia, and diagnostic services are included. A small number of physician services are not insured (e.g., cosmetic surgery, examinations for employment or insurance).

Data collected are based primarily on physicians' claims for payment on a fee-for-service basis. There are also physicians on alternative payment arrangements (e.g., salary, contract). Physicians on alternative payment may submit shadow or dummy claims, however not all services provided may be captured consistently. (In 2002–2003, these non-fee-for-service funding arrangements accounted for about 26%

of expenditures for physician-delivered services.) Claims data are accessible electronically from 1971 to the present; however, for practical purposes, data are typically used from 1975 forward.

Records include patient identifiers, details regarding the service provided, one diagnosis, and payment information. Diagnoses are recorded using three-digit ICD-9 codes (ICD-8 coding was used in data before 1979) and about 40 diagnostic codes assigned by Saskatchewan Health. Procedures are coded using fee-for-service codes from a payment schedule established through consultation between Saskatchewan Health and the Saskatchewan Medical Association. Because diagnostic data are given only to support the claim for payment and because only one three-digit ICD-9 code is recorded per visit, health outcome studies should not be done with unvalidated physician services data alone. When used in conjunction with other databases, however, the physician services data can play a useful role in data linkage projects.

Cancer Services Data

The Saskatchewan Cancer Control Program encompasses prevention, early detection, diagnosis, treatment, and follow-up of patients as well as research and education for malignant or pre-malignant disease. Provincial legislation mandates that information from medical professionals and hospital records required to complete the cancer registration must be provided to the Saskatchewan Cancer Agency. Thus, the cancer registry has a record of all people in the province diagnosed with cancer. In situ cancers and some neoplasms of uncertain behavior are also registered and followed. Within Canada, patients who move out of the province receive continued surveillance. All cases of invasive cancer are maintained in a follow-up program for a minimum of 10 years. The rate of loss-to-follow-up is approximately 3%.

This population-based registry was established in 1932. Complete computerized data for all cancer sites are available since 1967. For research purposes, data are usually used only from 1970 forward. The cancer registry contains identification, case, death, and review information and can be related to radiotherapy and in-clinic chemotherapy treatment data. Both confirmed and suspected cases of cancer are registered.

Vital Statistics

All birth, death, stillbirth, and marriage data are collected by Saskatchewan Health. Electronic data are readily accessible from 1979 to the present. Cause of death is recorded

by a physician or coroner on a Medical Certificate of Death form. The causes of death recorded on this form are keyed electronically and an algorithm is applied to determine the underlying cause of death in accordance with World Health Organization criteria. For events occurring up to and including 1999, four-digit ICD-9 codes are used to report cause(s) of death; since January 2000, coding is based on the ICD-10.

Live birth registrations record obstetric information and infant information. Completion of the Live Birth Registration form is the responsibility of the family. Although health information regarding the infant is not captured on the birth registration form, some information regarding the health of the infant, especially major congenital anomalies, may be found in the hospital services database because most births (over 99%) occur in a hospital. Stillbirth registrations include a "medical certificate" section, which is completed and signed by a physician or coroner.

Other Saskatchewan Health Information

Sundry health services data (supportive care services such as long-term care and home care services, some publicly funded mental health services, and laboratory services provided by the Saskatchewan Health Provincial Laboratory) are available and can be either linked with the HSN or manually reviewed to provide additional information on study populations. The tenure and completeness of the information in these databases are variable, and the suitability of these data for research would be dependent on the particular project.

Medical Records

Hospital record abstraction has been used for various studies to collect additional information to complement or validate information derived from the administrative databases. Medical records in hospitals are accessible upon approval from individual health authorities and affiliated facilities following established policies and procedures. Record retrieval rates have been excellent and typically exceed 95%.

Primary health records held by physicians have been accessed for research on several occasions, but retrieval rates are considerably lower than those achieved with institution-held records.

STRENGTHS

One of the greatest advantages of Saskatchewan's health databases is the use of the unique HSN to identify individuals. This number is used to code all health care services for an individual and hence can be used to link data from any of the computerized databases.

With health services data recorded for nearly the entire Saskatchewan population of one million, this is a population-based database. The registry is dynamic and updated daily; therefore it is very useful for providing current, valid denominator data. The database captures most prescription drug use in the province.

Data housed within Saskatchewan Health are electronically linkable and information can be compiled across the databases and over time. By linking data from two or more databases, it is possible to carry out both cross-sectional and longitudinal outcome studies. Given the long tenure of the databases, it is possible to compile information about prior drug use and previous disease experience for study populations.

The hospital separation and physician services data use standard international coding systems, enabling researchers to compare information from Saskatchewan with that from other jurisdictions. Access to hospital medical records is possible and retrieval rates have been excellent.

For pharmacoepidemiologic studies, it is important that data validity and reliability be evaluated. The various claims processing systems have built-in audit and eligibility checks. For research purposes, however, further validation is necessary. Validation, mostly by hospital chart review, has been built into several studies using Saskatchewan data. The validity of the diagnostic data is generally very good, but depends on a diagnostic code properly representing the condition in question and therefore varies with the condition. Hence, validity should be quantified for each condition studied.

WEAKNESSES

The health databases have been constructed by the Saskatchewan government primarily for program management purposes. Research is a secondary use, and the databases may not be well suited to some types of studies. Changes in policy and/or program features may influence the data collected in these mainly administrative databases.

The current population of Saskatchewan is relatively small for the evaluation of rare risks. To some extent, this limitation is mitigated by the fact that almost 30 years of drug exposure and outcome/diagnostic data are available. Nevertheless, the databases still may be too small to evaluate low prevalence exposures or rare outcomes.

There are some limitations regarding exposure data. The Drug Plan operates on a formulary system. While the Saskatchewan Formulary is extensive, the drug must be covered by the Drug Plan for records of its use to be included in the drug database. Also, there is no centralized computer database on inpatient drug use, OTC drug use, or use of alternative therapies. A new system of collecting data by

Saskatchewan Health on all prescription drug use by all residents of Saskatchewan begun in the summer of 2005 may enable research on drug use by Saskatchewan residents as soon as the drugs are first marketed in Canada. This would be valuable to enhance the capability of pharmacoepidemiologic studies that would not be limited by program coverage criteria.

Diagnostic information is derived primarily from hospital separation or physician billing data. If the outcome does not result in any medical attention, it cannot be identified. Also, if the outcome does not result in hospitalization, the diagnostic information is weaker because it is based on physician billing data, which have less complete and less specific diagnostic coding. The lack of a complete, centralized laboratory information database limits the ability to detect outcomes that must be identified and/or confirmed by specific test results.

In addition, the databases do not contain information on some potentially important confounding variables (e.g., smoking, alcohol use, occupation, and family history). Study designs must consider alternative methods of obtaining the necessary information on confounding factors (e.g., use of medical records or patient surveys).

Some of these limitations may be offset by the ability to access individual patient records to obtain information not included in the computer databases. (See Case Example 12.6.)

CASE EXAMPLE 12.6: STATIN USE AND RISK OF BREAST CANCER

Background

- The Cholesterol and Recurrent Events (CARE) study, a randomized controlled trial (RCT) of pravastatin in the secondary prevention of coronary heart disease, found a statistically significant increase in breast cancer risk among postmenopausal women treated with pravastatin for 4–6.2 years when compared with placebo.
- Pravastatin is one of the “statin” group of drugs used to treat hyperlipidemia.
- Most other studies of statin drugs lacked the power and duration to study breast cancer risk.

Issue

- Using the Saskatchewan administrative databases, a population-based study was designed to investigate a possible association between statin use and breast cancer.

Approach

- A historical cohort design.
- Female statin users during the study period (1989 to mid-1997) and an age–sex-matched comparison group of subjects not exposed to lipid-lowering drugs during the study period were selected.
- Subjects were followed forward to the study end date, defined as the earliest of the following: date of breast cancer diagnosis, date of death, coverage termination, or the end of the study period.
- The effects of age, exposure time, prior use of hormone replacement therapy (HRT), and contacts with physicians were considered.

Results

- 13 592 statin users and 53 880 non-exposed subjects were identified.
- Among women aged 55 years or younger, statin use was not associated with breast cancer incidence.
- In women older than 55 years, the relative risk of breast cancer was 1.15 (95% confidence interval [CI] 0.97–1.37) and stratified analyses revealed an increase in breast cancer risk in short-term statin users. An increase in breast cancer risk was also seen in statin users with long-term (>6 years) HRT use (relative risk = 2.04; 95% CI 1.20–3.46).

Strengths

- Population-based data enabled identification of most statin users in the province so that a much larger number of female statin users were included than in other studies.
- Reasonably long follow-up period of up to 8.5 years.
- Use of the cancer registry allowed for complete ascertainment of cancer cases.

Limitations

- Exposure to statins was estimated based on administrative records of dispensed prescriptions. This assumes that dispensed prescriptions were consumed.
- The number of cases was not sufficiently large to examine breast cancer risk for each of the statin drugs individually.
- Administrative databases lack information on some potential confounders.

(Continued)

Summary Points

- Observational data from administrative databases have a role in investigating drug effects found in RCTs in a “real world” setting.
- Chance and potential confounding cannot be ruled out as possible explanations for the slight increase in overall risk observed.
- More studies are needed to further investigate the observed differential risk associated with length of exposure to statins and HRT.

AUTOMATED PHARMACY RECORD LINKAGE IN THE NETHERLANDS

The Dutch system of medicine is based on primary care physicians—general practitioners (GPs) called “house doctors”—who practice in the community but not in hospitals, referring ambulatory patients to specialists for out- or in-patient care, as circumstances require. Hospital care is provided by full-time staff physicians representing various specialties. Medical care, including prescription drugs, is essentially paid fully by public or private insurers. Health care of virtually uniform quality is provided to all citizens. Hospitals are organized regionally, with well-defined catchment areas. It is, in principle, possible to link data from the community and hospitals, although with strict attention to anonymization. The health insurance system provides essentially complete insurance coverage for the total population.

DESCRIPTION

General Practice System

Since the late 1970s, numerous (mainly) academically-driven research networks of “sentinels” in general practice have been established among primary care physicians conducting population-based pharmacoepidemiology studies. A significant development in making use of the GP as a resource for pharmacoepidemiology has been the establishment of the Integrated Primary Care Information (IPCI) system, which is a research-oriented database with information from computerized patient records of GPs throughout the Netherlands. The database includes all demographic information, patient complaints, symptoms and diagnosis (International Classification of Primary Care, ICD-10), laboratory tests, discharge and consultant letters, and detailed prescription information (drug name, ATC code, dosing information and indication). The system maintains a

network of 150 general practitioners, covers over 500 000 people, is population based, and provides a potent tagging system enabling prospective follow-up of clinically well-defined target groups. The database has also shown its strengths for conducting a wide array of population-based pharmacoepidemiologic studies.

The Dutch Community Pharmacy System

Dutch community pharmacies are typically three to four times larger than their counterparts in other western European countries or North America, having 8000–10 000 patients per pharmacy. Dutch pharmacies essentially limit themselves to prescription drugs. They carry OTC products, but keep them behind the counter. Computerization of pharmacy records, and thus the compilation of prescription drug histories, is almost universal. Although recent public policies on pharmaceutical services aim to encourage patients to seek the most cost-effective pharmaceutical care available, and if necessary to switch from one pharmacy to another, practice and research show that mobility between pharmacies, which would hamper continuity and completeness of prescription drug histories, is virtually nonexistent.

Grouping patients as recipients of a particular pharmaceutical is thus possible by linking medication files, creating pharmacy-based cohorts of recipients of prescription drugs.

Medical Record Linkage

There is widespread, deep-seated resistance in the Netherlands to using unique personal identification numbers. Identification for insurance purposes is done by an anonymized family number, with substring codes used to distinguish individuals within the family and enable linkage of individual patient's records. To address the occasional need to ascertain specific information from an individual patient's medical records, in several studies community pharmacists have acted as an intermediary between the epidemiology researcher and the responsible physician.

In the early 1990s, a formal system of record linkage for pharmacoepidemiology was developed, called PHARMO. Developed by Herings and Stricker, it links community pharmacy and hospital data within established hospital catchment regions, on the basis of patient's birth date, gender, and GP code, preserving anonymity. While there are certain probabilistic aspects of the linkage, the combination of these three data items yields a sensitivity and specificity of over 95%. The PHARMO system has been expanded to a population of over 500 000, is population based, links all prescription drug data to hospital data, and has been used to study numerous drug effects severe enough to require hospitalization. PHARMO has linked also to primary care data,

population surveys, laboratory and genetic data, cancer and accident registries, mortality data, and economic outcomes. Even in the absence of a unique patient identifier—which remains missing in the Netherlands—the PHARMO system provides a powerful approach to conducting follow-up studies, case–control studies, and other analytical epidemiologic studies for evaluating drug-induced effects. The data collection is longitudinal and dates back to 1987. The system has well-defined denominator information, allowing incidence and prevalence estimates, and is relatively cheap because existing databases are used and linked.

Two more recent developments in Dutch pharmacy record linkage are the Rotterdam Study and the Prevention of Renal and Vascular End-stage Disease (PREVEND) study, both potent resources for pharmacoepidemiologic research. The Rotterdam Study started in 1990 as a population-based prospective follow-up study. All 10 275 residents of the Ommoord suburb in Rotterdam aged 55 years or over were invited to participate. The baseline measurements include a physical examination, demographic data, medical history, family history of diseases, and lifestyle factors. Moreover, blood samples are drawn for DNA extraction. Pharmacy records have been linked to this cohort since 1991, providing a powerful resource for studying time-dependent drug effects.

PREVEND was initiated by Groningen University and based on a cohort of patients with micro-albuminuria identified in the general population. This group and an equally sized sample of non-microalbuminuria subjects are being followed for various outcomes, and linkage to pharmacy records has been achieved in order to conduct pharmacoepidemiologic research.

STRENGTHS

The most important advantage of the Dutch system lies in its virtually complete coverage of a relatively homogeneous population of reasonable size. The quality of data on drug exposure in the Netherlands is excellent because of three factors:

1. computerized dispensing records are subject to financial audit because they are the basis for reimbursement;
2. a long tradition that patients frequent a single GP and pharmacy;
3. the practical lack of economic incentives for between-pharmacy shopping.

Data on drug exposure can be linked extensively to various outcomes and to data on possible confounders or effect modifiers. (See Case Example 12.7.) With respect to the

latter, the strong increase in possibilities of linking genetic information to studies evaluating drug exposure–outcome associations is an important development.

Yet, it is not only genetics that drives more focus on molecular pharmacoepidemiology in the Netherlands. Researchers studying the nature and extent of drug effects have found their way to laboratory medicine and other resources where biochemical and other clinical “signatures” of patient outcomes can be ascertained.

WEAKNESSES

Unreliable or outdated information in the patient file has occasionally undermined the quality of pharmacy records. Patient files in pharmacies are maintained mostly by pharmacists, on the basis of information provided by patients, relatives of patients, or local administrative sources. This eclectic sourcing of information limits the value of demographic information contained in pharmacy computers.

There is an ongoing need for quality assessment of data for epidemiology research derived from medical information systems. There is thus a need for careful monitoring of the quality of medical registries, including the process of data recording, its completeness, and its validity.

A second weakness is the size of the country, and thus the population exposed to certain drugs, together with the general reluctance within the medical community, especially general practice, to adopt new drugs after marketing approval. Rare events, particularly when exposure is small, are difficult to study. Probably the biggest potential risks to the future of the Dutch pharmacy database are the consequences of politically motivated tinkering with a health care system that has worked very effectively and, relative to other countries, economically.

CASE EXAMPLE 12.7: **TIME-DEPENDENT DRUG EFFECTS**

Background

- Since the mid-1990s there has been concern about the association between long-term use of calcium channel blockers (CCBs) and cancer.

Issue

- Earlier findings of association between CCBs and cancer are possibly confounded by the use of cross-sectional data on drug exposure.

(Continued)

Approach

- Prospective population-based cohort study. Time-dependent analyses of CCB exposure in relation to incident cancers between 1991 and 1999.

Results

- No associations between CCBs and cancer could be confirmed except for verapamil.

Strength

- Prospective cohort study with access to very detailed data on time-dependent drug exposure over time.

Limitations

- Occurrence of cancer is multifactorial and not all relevant determinants are known and measured.

Summary Points

- Cancer as an outcome in pharmacoepidemiology requires long periods of follow-up.
- Time-varying drug exposure requires time-dependent analyses.
- Be cautious about class effects.

TAYSIDE MEDICINES MONITORING UNIT (MEMO)

The National Health Service in Scotland (NHSiS) is a tax-funded, free-at-the-point-of-consumption, cradle-to-grave service. In Scotland, there is very little private health care and as there are no socioeconomic eligibility distinctions the level of individual health care is based on need alone.

The Medicines Monitoring Unit (MEMO) is a university-based organization that works closely with the NHSiS to record-link health care data sets for the purposes of carrying out research. MEMO was founded in the late 1980s to perform in-hospital studies of adverse reactions. However, because most prescribing occurs in the community, in the 1990s the research direction changed towards community-based studies. Since then the MEMO has conducted many studies to detect and quantify serious drug toxicity in the community, using record-linkage techniques. However, MEMO has also diversified its activities to undertake outcomes research and economic, genetic,

drug utilization, and disease epidemiologic studies. MEMO has traditionally used data from the Tayside region of Scotland, which is geographically compact and serves over 400 000 patients. It has a low rate of patient migration. Health care for the region is coordinated by NHS Tayside (www.nhstayside.scot.nhs.uk), part of the NHSiS. NHS Tayside maintains a computerized record of all patients registered with general practitioners. Additionally, Tayside has been at the forefront of generating managed clinical networks for chronic diseases such as diabetes, heart disease, and endocrine disorders.

At the heart of the MEMO record-linkage system is the dispensed prescribing data set, which is unique to the UK. In the absence of data on whether patients actually take their medications as directed, this is the acknowledged prime data set in which to establish exposure and drug utilization patterns. There are other key electronic data sets regularly used by MEMO. The most important of these are the national Scottish Morbidity Record (SMR) databases (www.statistics.gov.uk/STATBASE/) that comprise the electronic health records for Scotland. Additional databases include cancer registration, community dental services, cardiac surgery, and drug misuse databases. In addition, there is the Scottish Immunisation Recall System (SIRS), which tracks all childhood immunization in Scotland.

MEMO makes use of Tayside regional biochemistry, pathology, hematology, microbiology, and hospital clinic data as well as data sets imported or created specifically for research projects. MEMO also now uses Scotland-wide data, based on record-linkage using patients' names, dates of birth, and addresses. It has mainly been used for evaluation of health care activity and outcomes research. In conjunction with MEMO, the ability to link dispensed prescribing data to hospitalizations for Scotland's five million people has been demonstrated. This allows the evaluation of drug safety and accurate quantification of risk of adverse drug reactions with less commonly prescribed drugs. Such work has shown the utility of the record-linkable Scottish unique identifier, and as a result about 85% of all prescriptions now contain this identifier. This now enables large-scale cost-efficient database tracking studies that may become the norm in future pharmacovigilance and outcomes research. Additionally, the managed care networks already operating in Tayside are now being rolled out over all of Scotland.

Importantly, whether working with Tayside or Scottish data, MEMO has been at the forefront of ensuring compliance with the principles of good epidemiologic practice (GEP) with regard to the use of observational data.

DESCRIPTION

Patient Identification

Every person registered with a general medical practitioner (GP) in Scotland is allocated a unique identifying number known as the Community Health Index number (CHI number). This is a ten-digit integer, the first six digits indicating the date of birth, digits seven and eight giving region of residence information, the ninth digit indicating sex, and the tenth digit incorporating a checksum to ensure the validity of the number. The CHI number, which serves as patient identifier, maps to a data set that serves as a useful roster file in epidemiologic studies. For any data with just a name and address, MEMO has powerful software that enables the CHI to be added.

For practical purposes, the entire Tayside population is registered with a GP and thus appears in the central computerized records held by the Health Board, the Community Health Master Patient Index. This file also contains GP and address details, and a log of deceased persons along with dates of death. The demographic composition of the Tayside population can therefore be readily obtained.

Dispensed Prescription Drug Data

In Scotland, all community prescribing is performed by GPs, sometimes on the advice of hospital physicians. Only hospital inpatients receive drugs by a different mechanism. MEMO has devised a method of capturing all GP prescribing. Upon receipt of a prescription by a GP, the patient takes it to any of the community pharmacies where the prescription is dispensed. The pharmacist then sends the original prescription form to the Practitioner Services Division (PSD) of the ISD of the NHSiS in Edinburgh to obtain reimbursement. After paying the pharmacists and dealing with any appeals, the PSD sends the cashed prescription forms to MEMO, where they are stored on a database linked to the CHI number.

Ascribing the Unique Identifier Number to Prescriptions

All prescriptions are electronically scanned and read by powerful optical character recognition (OCR) software that runs multiple algorithms to obtain the best recognition for differing type faces and font sizes. Each area of the prescription is read against a dictionary of the text that might be expected (name, address, drug name, dose instruction, doctor name, date, pharmacy name) to generate an OCR best recognition. Both the actual text scanned and the OCR

recognition are stored in the database. The objective at this stage is to "sort" all prescriptions by drug items so that data entry can concentrate on prescriptions by drug name or BNF code, making the data entry more time efficient.

The next stage involves the reconciliation of all data not picked up accurately by the OCR system. All prescriptions are viewed electronically on screen by a data entry clerk, who validates data and enters handwritten prescriptions that the OCR has failed to read. Another data entry operation then adds the CHI number to the prescriptions where this is missing (about 15% of prescriptions), and verifies the dosing instructions.

MEMO now receives all prescriptions electronically as scanned images. With each image file is a text file that consists of details of the prescription drug, the cost, the prescribing doctor code, the practice code, and (in 75–80%) the CHI number. MEMO uses these data sources to create a dispensed drug file that has complete CHI ascribed and also has dose and duration of treatment added. This solution is scaleable to Scotland.

Prescriber Data

The code number of the GP with whom the patient is registered is known and the code number of the GP issuing the script is held with each prescription record. Repeat prescriptions may be written by GP partners in rotation. This method of data collection allows the identification of the prescriber who initiated a treatment course.

Hospital Data

Since 1961, all hospitals in Scotland have been required to compile and return coded information on all acute inpatient admissions, forming the basis of the Scottish Morbidity Record 1 (SMR1), which contains administrative, demographic, and diagnostic information. Each SMR1 record has one principal and five other diagnostic fields coded according to the ICD-10. There is also one main operation or procedure field and three others coded according to the Office of Population and Census Surveys, 4th revision (OPCS4) classification. In Tayside, there are approximately 63 000 hospital discharges per year. MEMO has in-house historical SMR1 data dating back to 1980. These data allow for a past medical history of hospitalization for a condition to be controlled for in data analyses.

The SMR1 database contains details of deaths certified in hospitals, which may be up to 85% of the total mortality. The Community Health Master Patient Index records the

date of death of subjects in the Tayside population, while information on the certified cause of death of patients is provided to MEMO by the Registrar General. Population-based morbidity and mortality studies are thus feasible.

Other In-Hospital Data

Any health care data set that is indexed by the CHI number can be linked into MEMO's record-linkage database. In MEMO, commonly used data sets are the cancer registration database, child development records, maternity records, psychiatric records, and neonatal discharges.

Clinical Laboratory Data

Clinical laboratory data for the Tayside region since 1989 are held on a computerized archive in the Department of Biochemical Medicine in Ninewells Hospital. The database has CHI-specific biochemical, hematology, microbiology, virology, and serology laboratory results and reports. CHI-specific results from all pathology investigations since 1990 for Tayside are stored electronically in MEMO. Those data can be record-linked to the MEMO database to complete the clinical characteristics of disease or hospital admission. It also allows investigators to study the effectiveness of drug treatments such as lipid-lowering drugs and anti-diabetic drugs by monitoring serum lipid profile, serum glucose level, glycosylated hemoglobin levels, etc.

Primary Care Data

Health care data stored on GP computer systems are increasingly available. At present, such linkages are done *ad hoc* for each study, but this is becoming increasingly integrated using new purpose-built data extraction software. Access to data such as smoking, alcohol use, body mass index, blood pressure, visits to GPs and nurses, as well as access to symptoms and diagnoses are possible. MEMO also has a bank of research nurses who extract data written in paper case records, providing a methodology for accessing all GP data in a patient's record since birth.

Geographic Data

All patients and their addresses are known, including post code, and information is available from the decennial census regarding the relative deprivation levels of small post code areas in the form of the Carstairs deprivation score. This is a z-score using the following census variables: unemployment, overcrowding, male car ownership, and low social class. The Carstairs deprivation score can be used as an indicator of the socioeconomic status of patients.

Studies of geographic variation in prescribing and outcome are also feasible. Post code data also allow the merging of other types of environmental data, including climatic and meteorological, into the data set.

DARTS and Hearts Managed Clinical Network Databases

MEMO has facilitated the development of clinical systems to promote seamless care of patients across all parts of the health care system; as such all health care professionals use just one information technology system. The Diabetes Audit and Research in Tayside Scotland (DARTS) database is the most developed of these resources and essentially is record-linkage, at source, in real time at the point of care delivery. These databases provide rich and detailed data on the disease phenotype, severity, complications, associated comorbidities, non-drug treatments, demographics, etc. The DARTS system has now been adopted nationally in Scotland and is known as SCI-DC (www.show.scot.nhs.uk/crag).

Birth Cohort 1952–1967 (Walker Data Set)

This birth cohort, named after the professor of obstetrics who created it, is a database of more than 48 000 birth records containing meticulously recorded details of pregnancy, labor, birth, and care before discharge for babies born in the Dundee area in the 1950s. MEMO has been able to add the CHI to over 21 000 children, presently alive and living in Tayside (now aged 37–51), and additionally over 15 000 mothers and fathers. Linkage across siblings and over two and three generations is now possible. This data set forms a powerful resource for familial genetic studies by linking with the phenotypic data available within the other data sets, with intermediate phenotypic data obtained at the time of sample collection as well as future database tracking.

Ambulance and Emergency Department Data

MEMO has the ability to obtain data on ambulance use and emergency visits to the Accident and Emergency departments. Such data have enabled MEMO to provide a complete picture of all resources used by diabetic patients who have hypoglycemic attacks that require NHS hospital treatment but who do not require hospitalization.

Other Outcome Data Sets

Other health care data sets are available in Tayside that are not indexed by CHI number. However, provided that some patient demographic details are present, such as name, date of birth, and post code, MEMO can usually identify the correct CHI number in the same way that CHI

numbers are identified for prescriptions. Thus, biochemical laboratory reports filed on computer tape back to 1977 and computerized records of histopathology reports have been used. MEMO has also constructed a database of 100 000 endoscopy and colonoscopy procedures, and, in collaboration with Tayside Police, a database of subjects involved in 22 000 road traffic accidents in Tayside.

Patient-Reported Outcomes

Over the past three years, MEMO has conducted several studies involving direct contact with patients to obtain quality-of-life information, patient-reported outcomes such as diabetes-related hypoglycemic attacks managed at home, or the effects of asthma on day- and night-time activities. More recent work is aimed at understanding factors that lead to non-adherence to medications.

Genetic Data–Phenotypic Linkage

The linkage of genetic information with phenotypic data is a task being undertaken in many parts of the world. The molecular biology techniques used to detect and analyze genetic information are now fairly routine, having been enabled by increasingly powerful and rapid technological advances. However, the phenotypic component is the area that deserves most care as linkage of genetic data to phenotypic data of unknown, poor, or mediocre quality is at best poor science and at worst likely to miss important associations. The quality of Tayside data as recorded in the various data sets, the coverage of the data sets, the cradle-to-grave single data collection point nature, and the new managed care clinical networks produce phenotypic data of high quality, enabling pharmacogenetics and pharmacogenomic research.

Confidentiality, Ethics, and Good Epidemiologic Practice

European and UK data protection law, NHS guidelines, and research governance requirements, as well as published codes of GEP, clearly dictate what can and cannot be done with UK health care data as well as defining how such data should be managed in observational studies. The tight geographic nature of data used by MEMO as well as the fact that the data come from one NHSiS region, along with the climate of “nervousness” that surrounds data linkage, particularly for genetic purposes, have meant that MEMO has had to undergo a radical evolution to allow it to continue doing pharmacoepidemiology and other research. This evolution has seen MEMO split into three separate organizations over

three physical locations in order to separate subject identifiable data from researchers. In addition, MEMO now undergoes external audit of all activities, the project management and external audit having as their main focus adherence to GEP.

The three separate organizations include:

1. The NHS Clinical Technology Centre, which stores and processes all NHS subject-specific data. No research is undertaken by this group.
2. The MEMO/HIC/NHS Clinical Interface Unit, which acts as a research “bureau service” in that it provides anonymous record-linked data for research. The unit takes in data with patients’ names and addresses and exchanges these for the CHI number and then, because the CHI is a personal identifier (containing sex and date of birth identifiers), exchanges the CHI for an anonymized number (ANO-CHI). A further anonymization is then undertaken before release of data for a research project. This final anonymization ensures that data from separate studies cannot be linked without referral back to the ANO-CHI level, and this can only be done against strictly controlled approved protocols under the guidance of the Tayside Caldicott Guardians (the government-appointed guardians of subject identifiable health data).
3. The Health Informatics Centre (HIC), which is a new custom-built center that houses the research and biostatistics part of MEMO.

STRENGTHS

Patient Identification

One of the greatest advantages of using data from Tayside is the unique patient identifier. This allows for relative ease of record-linkage and, since this number is also age- and sex-specific, it is relatively easy to choose age-, sex-, GP-, or practice-matched comparator groups from the population. Selection of patients for both cohort and case–control studies is thus efficient.

Population-Based Data

MEMO is regularly supplied with updated copies of the Community Health Master Patient Index from NHS Tayside, and uses this to track the population of patients alive and residing in Tayside on a daily basis to define study populations for drug safety studies. Unlike clinical trials, which focus on highly selected patients, this observational approach allows “real-world” populations to be studied representing all socioeconomic groups and within

a universal health care coverage scheme. Such population-based data allow the calculation of incidence rates, excess risk, and attributable risk.

Drug Exposure Data

The data captured at MEMO represent prescriptions dispensed at a pharmacy; therefore, noncompliance with filling prescriptions is eliminated. This rich resource allows researchers to carry out a wide variety of population-based research. (See Case Example 12.8.)

Accessibility to Medical Records

A major strength of MEMO is the ability to examine original hospital case records. This allows for quality control of the computerized data and can also deal with some elements of confounding. Over-the-counter (OTC) preparations are not recorded in the MEMO drug exposure database and this confounding can be partly controlled by such data. Similarly, information may be available on potential confounding factors such as smoking and alcohol consumption.

High Diagnostic Accuracy

Record-linkage of data sets using a diagnostic algorithm rather than the use of a single diagnosis code from one data set improves the specificity and sensitivity of diagnoses. In addition, the access to case medical records further improves accuracy.

Low/No Missing Data

By using data from computer systems, MEMO data are not subject to missing data on hospitalizations or data that may have transcription errors. Historically, this has been a known problem for primary care data sets.

Patient Access

MEMO is able to use its data sets to generate lists of subjects who meet set criteria for entry to certain types of prospective studies. Such work is done under a detailed standard operating procedure ensuring patient confidentiality and in conjunction with the subject's primary care doctor, who is the point of contact with the patient.

Randomized Simplified Trials and Randomized Epidemiology

Access to patients, collaboration with primary care, and the MEMO data sets enable cost-efficient randomized Phase IV real-world studies of safety and/or effectiveness to be

done at low cost. These studies can contribute significantly to risk management plans after licensing. They can also be used to produce the data on effectiveness that are increasingly demanded by health service providers. In such studies, patients are recruited according to good clinical practice by their primary care physician. They are then randomized by MEMO to receive a study drug prescription or a comparator drug prescription, prescribed in the normal manner within the NHS. The relevant data sets are then tracked for exposure and outcomes. Once randomized, subjects are treated in the same way as other subjects. These trials are open but the endpoints can be assessed by a committee blinded to treatment allocation—the so-called PROBE (Prospective, Randomized, Open, Blinded Endpoint) design. Importantly, those randomized into such studies can be compared with those who are not randomized to determine if those randomized are representative of the real-world population who get such drug treatment.

WEAKNESSES

Population Size and Drug Use

The current population of Tayside is approximately 400 000, which is comparatively small for many pharmacovigilance studies. However, 400 000 is an adequate size for many other types of pharmacoepidemiology studies.

Drug exposure data in Tayside are only available from 1989 and cover only a limited set of drugs until January 1993 (from when all dispensed prescriptions have been collected).

OTC and Hospital-Dispensed Medications

Another weakness is the inability to capture directly exposure to OTC drugs or drugs prescribed in hospitals. People use OTC medications such as aspirin, ibuprofen, paracetamol (acetaminophen), and others for their symptoms.

Indication for Use

Given that confounding by indication is arguably one of the most difficult potential sources of error in pharmacoepidemiologic research, one of MEMO's biggest weaknesses is that the diagnostic indication for prescribing is not available. Where the indication for a drug is wide, this can lead to difficulties. In addition, there is a certain amount of "over-coding" that happens in primary care data sets where the coded diagnosis given at the time of prescription is assigned for convenience and may be incorrect.

Inaccurate Diagnosis

One of the criticisms leveled at record-linkage studies is the inaccuracy of computerized medical diagnoses. In MEMO, the discharge diagnoses for SMR1 are abstracted from the clinical discharge summaries by specially trained coding clerks. These clerks occasionally have to interpret “soft diagnoses,” such as symptoms for which no cause can be found. In addition, nonstandard terminology may be employed to describe an illness, so the coding of diagnoses may be imprecise. Computerized algorithms exist to detect and reject the most glaring errors, but errors of interpretation persist within the database. Nevertheless, these SMR databases are continually audited by the ISD of the NHSiS and the quality and accuracy of the data they contain are consistently high.

Episodes and Admission

Another issue that must be appreciated with computerized SMR data is that the currency is the “consultant episode of care.” In other words, a patient who is hospitalized with a gastrointestinal hemorrhage under the care of an internist and who then gets transferred to the care of a surgeon, gets transferred to an intensive care specialist postoperatively, and then back to an internist before discharge will have four computerized admission and discharge records, each containing diagnostic terms and procedure terms. Government statistics on health care activity, often used in determining power calculations for drug safety studies, list these consultant episodes and not hospital admission events. Typically, 100 consultant episodes would translate into about 70 admission and discharge events. The weakness of this system is that consultant episodes must be reconciled into admission and discharge events. The strength is that more diagnostic terms are used and the likelihood of incorrect coding is minimized. In addition, a major benefit of these episodes of care data is the ability to “cost” hospitalizations. Each facility the patient is treated in carries a different NHS tariff cost. The economic evaluation of health care interventions is assuming increasing importance within any health care system so these data are of increasing importance.

CASE EXAMPLE 12.8: **GLUCOCORTICOIDS AND CARDIOVASCULAR** **DISEASE**

Background

- Glucocorticoids have adverse systemic effects, including obesity, hypertension, and hyperglycemia, that may predispose to cardiovascular disease.

The effect of glucocorticoid use on cardiovascular disease has not been quantified.

Question

- Do users of exogenous glucocorticoids have an increased risk for cardiovascular disease?

Approach

- A population-based cohort study using a record linkage database.
- 68 781 glucocorticoid users and 82 202 nonusers without previous hospitalization for cardiovascular disease were studied between 1993 and 1996.
- Measurements: the average daily dose of glucocorticoid exposure during follow-up was categorized as low (inhaled, nasal, and topical only), medium (oral, rectal, or parenteral <7.5 mg of prednisolone equivalent), or high (≥ 7.5 mg of prednisolone equivalent).
- Poisson regression model, sensitivity analysis, and propensity score methods were used to investigate the association between glucocorticoid exposure and cardiovascular outcome.

Results

- 4383 cardiovascular events occurred in 257 487 person-years of follow-up for a rate of 17.0 (95% CI 16.5–17.5) per 1000 person-years in the comparator group, and 5068 events occurred in 212 287 person-years for a rate of 23.9 (95% CI 23.2–24.5) per 1000 person-years in the group exposed to glucocorticoids (22.1, 27.2, and 76.5 in low, medium, and high groups, respectively). The absolute risk difference was 6.9 (95% CI 6.0–7.7) per 1000 person-years (5.1, 10.1, and 59.4, respectively).
- After adjustment for known covariates, the relative risk for a cardiovascular event in patients receiving high-dose glucocorticoids was 2.56 (95% CI 2.18–2.99).

Strengths

- Population-based cohort design with complete follow-up over the study period.
- Advanced methodology such as propensity scores and sensitivity analyses to reduce and control the bias and unmeasured confounding

(Continued)

Limitations

- The study did not definitively determine that glucocorticoids themselves, not the diseases that had necessitated treatment with high-dose glucocorticoid treatment, accounted for the higher risk for cardiovascular disease.

Summary Points

- Treatment with high-dose glucocorticoids seemed to be associated with increased risk for cardiovascular disease.
- Lower doses were not associated with an increase in cardiovascular events.
- Patients and doctors should consider this potential risk when they weigh the benefits and risks of glucocorticoid treatment.

UK GENERAL PRACTICE RESEARCH DATABASE

Databases that collect 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 on their patients. Administrative databases are often inferior to medical records databases when studying the presence or absence of a disease because their intended purpose is often for billing. Consequently, they often fail to capture important health data (e.g., family history, lifestyle practices) and the health care data that are captured may suffer from significant validity limitations in that their purpose was for the generation of a bill or payment and not patient care. In contrast, the purpose of a medical record database is to provide a health care provider with information needed to care for a patient. While on face value this information may seem to be of superior quality for epidemiologic studies, medical record databases can also have limitations, because the collection of data is for patient care and not necessarily for scientific studies. Therefore, the information recorded or even the way that data are quantified by the health care provider in the database may not reflect the interests of the epidemiologic investigator. The General Practice Research Database (GPRD) is generally considered to be the largest medical records database in routine use for epidemiologic investigation, and the one which has been used most extensively for published pharmacoepidemiologic research. (See Case Example 12.9.)

DESCRIPTION

History and Evolution of the GPRD

Initially called the Value Added Medical Products (VAMP) Research Databank, the GPRD was started in June 1987 as a tool for conducting public health research based on data routinely recorded by the GP using an electronic medical record. The system of health care delivery in the UK allows a unique opportunity for researchers to study data on a well-defined population. In the UK, virtually all patient care is coordinated by the general practitioner (GP) through the National Health Service (NHS). When patients are referred for specialty care, a treatment plan is initiated by the consultant but ultimately chronic therapies are prescribed and monitored by the GP. When patients are seen by specialists or in the hospital, future treatment is directed through the GP, ultimately allowing for capture of this information as well.

The database contains information on both medical diagnoses and prescription medications, which are recorded by the GP as part of the patient's medical record. The GPRD has been used internationally by researchers from academia, regulatory authorities, and industry.

Data Collection and Structure

In any given year, GPs contributing to the GPRD provide data on about 3 million patients, which translates into about 44.8 million person-years of follow-up between 1987 and 2006. Continuous information has been collected for six years or more in most of the practices, and over 1 million patients have more than 11 years of data. About 5% of the UK population is included in the GPRD, which is broadly representative of the general UK population in terms of age, sex, and geographic distributions. Since 1994, data have been collected by approximately 1500 GPs who work in 500 practices across the UK; these practices have agreed to record data in line with recording guidelines agreed between the GPRD and contributing GPs.

GPs use the computer software to enter data on their clinical encounters. These data are stored in different files: patient file containing demographic and registration information; clinical file containing medical coded information related to routine care or resulting from hospitalizations, consultations, or emergency care, along with the date and location (e.g., GP's office, hospital, consultant) of the event and an option for adding free text and including blood pressure measurements; referral and consultation files containing information on referral for specialist consultation and details of consultations test file containing

test results downloaded electronically from pathology laboratory; immunization and prescription contraceptive files; therapy file containing information on all issued acute and repeat prescriptions, including date of prescription, formulation, strength, quantity, and dosing instructions, indication for treatment for all new prescriptions (cross-referenced to medical events on the same date), and miscellaneous information such as smoking, height, weight, immunizations, pregnancy, birth, death, date entering the practice, date leaving the practice, and laboratory results.

Diagnoses were recorded using Oxford Medical Information System (OXMIS) codes until 1995 when the Read coding system was introduced. OXMIS codes are similar to ICD-9 codes; however, OXMIS codes allow for more detailed diagnostic coding. Read codes are alphanumeric codes that group and define illnesses using a hierarchical nosologic system. The Read codes are a very comprehensive coded clinical language developed in the UK and funded by the NHS. The codes include terms relating to observations (signs and symptoms), diagnoses, procedures, and laboratory and radiologic tests.

Prescriptions were originally entered using Prescription Pricing Authority (PPA) codes and are currently entered using Multilex codes issued by First Databank. Drug codes provide detailed information on the drug, dose, and route of administration. Patients in the GPRD are also issued a number that identifies people residing at the same address or not at the same address but who are family members. This data field, in combination with the date of birth of children and the delivery code in the mother's record, can be linked to identify mothers and their children.

Extensive studies have been performed to evaluate the validity and completeness of the GPRD. The validity of specialists' information and its capture by GPs in the GPRD have been well documented with studies demonstrating that 87% of diagnoses from specialist letters were documented electronically. There is also good agreement between GPRD prescribing data and national data from the PPA. Due to the UK's requirement for printed prescriptions and because the GP computer system provides this service for the GP, pharmacy prescribing information is very well documented in the GPRD.

The database, although extensive, may not contain data on every characteristic that may be required for a study. For example, information on occupation, employment, and socioeconomic status is not available electronically. Data recording has also been changing over time. Prior to about 2001 the quality of the recording was managed using GPRD Data Recording Guidelines, which meant that some events of minor significance might not have been recorded. Since

then, changes in the UK NHS have meant that there has been a fairly rapid movement to electronic notes and data moving electronically from hospital/laboratories to GPs over the NHS intranet-NHSnet. More recently the Quality Outcomes Framework (QOF) initiative has meant that GPs are rewarded for ensuring data on key diseases areas is well recorded.

STRENGTHS

Population-Based Data

As a data source, the GPRD provides researchers with the opportunity to design studies using population-based methods, which minimizes selection bias and improves the validity of epidemiologic studies. Population-based studies are defined as those in which the cases (e.g., individuals with disease) are a representative sample of all cases in a precisely defined population and the controls are directly sampled randomly from this population. The GPRD represents a defined population, which allows investigators to study all patients with a given disease and gives them the ability to study control patients from the same source population from which those with the disease of interest are derived. Therefore, even in conducting case-control studies, the GPRD offers a significant advantage for epidemiologic studies over traditional hospital-based designs. Furthermore, the GPRD is broadly representative of the UK population in general, suggesting that findings from the GPRD should generalize to the broader UK population.

The well-defined population of the GPRD also allows investigators to study families and to link health events in mothers to outcomes in their children. Finally, all individuals in the GPRD are assigned a practice number as well, allowing researchers to measure individual practice effects on health outcomes.

Size of the Database

Between 1991 and 1996 there was an average of 3.4 million active (i.e., not temporary) patients represented in the GPRD. The cumulative experience of the GPRD yields about 9.8 million patients followed for over 44.8 million person-years. As a result, the GPRD can be used by researchers to study rare outcomes using cohort designs for medical conditions with incidence rates of less than 1/10 000 with sufficient statistical power. The number of practices participating in the GPRD has been decreasing for the past five years as practice lists have become larger. However, the total number of active patients with usable lengths of medical history has been increasing and currently stands at over 3.1 million.

Validity of Information

The validity of the GPRD has been extensively studied. Studies have demonstrated good agreement between the electronic medical record and capture of information from specialists. Various individual diagnoses have been validated by direct query of the GPs, and the accuracy of information on prescription medications has also been demonstrated. Additionally, the quality of data submitted by GPs undergoes continual internal scrutiny by administrators of the GPRD. As a result, the GPRD is one of the best-studied sources of health data available for epidemiologic investigations.

Access to Original Medical Records

Investigators can obtain, through an intermediate, anonymized copies of the patient's (non-electronic) medical record and a more detailed review of the patient's health history. This capacity allows researchers to verify information captured on death certificates and letters from specialists. Studies in which medical records have been requested have resulted in response rates of over 80% and in many cases over 90%, with the majority of requests being met within three months. This service also allows investigators to send questionnaires to the GPs about individual patients. In some instances, researchers can have questionnaires completed by individual patients by working through the GP. All data from questionnaires or paper-based medical records are stripped of personally identifying information before being sent to the researcher. A new feature is the external record linkage of the GPRD to other NHS health care data sets such as the essentially complete central death record, hospitalization records, geographic variables and disease registries: this has important implications for both the validity and utility of research conducted in the GPRD.

WEAKNESSES

Completeness of Data

The GPRD is in most cases used by the GP as the patient's medical record, and therefore information generated by the GP is expected to be complete. However, information from hospitals may not have been fully captured in the electronic medical record prior to about 2002. Communication from specialists, discharge summaries from hospitals, and test results from pathology laboratories were, prior to 2002, often received in hard copy and manually entered into the computer. Data from hospitalizations or specialists may have been missing in about 10% of the records. In particular, minor medical events are more likely to be

missed than medically significant diagnoses or events. Information on treatments that are restricted by the NHS to specialist care (e.g., psoralen plus ultraviolet A therapy, cytotoxic/chemotherapy) may be particularly problematic. Laboratory data were also incomplete, analogously. Since 2002, increasing amounts of data are being entered electronically, and so completeness has improved. However, data on all medications and treatments given whilst in a hospital bed are not readily available. Data on non-significant medical events and exposures to medications that occurred before enrollment in the GPRD and are no longer active clinical issues may also not be documented in the electronic medical record. Use of the GPRD for more detailed diagnostic studies may be problematic. For example, investigators have determined that a diagnostic coding algorithm cannot reliably identify patients with pneumococcal pneumonia based on query of the GP, although investigators can reliably determine that a patient had pneumonia.

Complexity and Costs of Computer Hardware/Software Needed to Work with the GPRD

The size and complexity of the GPRD requires that those working with it have adequate computer software and hardware, as well as experienced data managers. Technical requirements for working with the GPRD will vary according to what version of the database is being used. GPRD is available through a web-enabled link, which requires knowledge of the particular application into which the data are loaded and stable telecommunications links, or as a set of flat text files that can be imported into any application but requires hardware and data storage facilities. GPRD raw data through March 2002 is also available through EPIC. Actual use of the GPRD may vary from user to user with respect to the size of the database, the fields of the database available for review, and whether data are available as individual records or tables.

CASE EXAMPLE 12.9: **THE RISK OF UPPER RESPIRATORY TRACT** **INFECTIONS IN ACNE PATIENTS EXPOSED** **TO CHRONIC ANTIBIOTIC THERAPY**

Background

- It has been hypothesized that long-term antibiotic use may lead to increased rates of infection through alteration in flora and the development of resistance.

- Acne patients are otherwise healthy, but frequently require chronic antibiotic therapy and therefore represent a natural population in which to test this hypothesis.

Question

- Does chronic (e.g. >6 weeks) use of oral antibiotics (erythromycin or tetracyclines) or topical antibiotics (erythromycin, clindamycin) in acne patients lead to an increased risk of upper respiratory tract infections?

Approach

- A cohort study using patients with acne from the UK General Practice Research Database (GPRD) 1987–2002.
- Exposed patients were those with acne who received at least 6 weeks of antibiotic therapy, and control patients were those with acne who did not receive antibiotics.
- A second unexposed group consisted of individuals with hypertension to determine if health care seeking behavior impacted the findings.
- Outcomes were any upper respiratory tract infection over a period of 12 months.
- The results were adjusted for sex, age, year of diagnosis, practice, number of prescriptions for acne antibiotics over the 12 months of observation, number of office visits for acne, history of diabetes, and history of asthma.

Results

- Acne patients exposed to chronic antibiotics had an increased risk (odds ratio 2.15, 95% CI 2.05–2.23) of developing an upper respiratory tract infection compared to acne patients not on chronic antibiotic therapy.
- The findings did not change when adjusting for potential confounders or measurement of health care seeking behavior.

Strengths

- This cohort study utilized a well-defined source population which minimizes bias.
- The large sample size allowed for adjustment for many potential confounding variables.
- The findings were robust to multiple sensitivity analyses.

Limitations

- Although the GPRD has been shown to capture the presence of an upper respiratory tract infection accurately, it does not accurately ascertain the source of the infection (e.g., viral, bacterial).
- The study may be limited by confounding by indication as only patients with moderate to severe acne are treated with chronic antibiotics and it is not known if acne patients are more or less prone to upper respiratory tract infections independent of antibiotic use.

Summary Points

- Chronic antibiotic therapy may increase the risk of upper respiratory tract infections—a potentially important public health finding as millions of patients are treated with antibiotics chronically for acne and other conditions.
- The GPRD has several advantages for conducting this study, including a large sample size, a well-defined source population, and data on prescription use and medical outcomes collected in a longitudinal manner.
- Additional studies, such as randomized controlled trials, are necessary to confirm the hypothesis as well as to better measure the impact of chronic antibiotic therapy on the source of upper respiratory tract infection.

CONCLUSIONS

This chapter describes nine existing medical databases that have been useful and productive in the conduct of pharmacoepidemiologic research. Included are the Group Health Cooperative, Kaiser Permanente Medical Care Program, HMO Research Network, UnitedHealth Group, Medicaid Databases, Health Services Databases in Saskatchewan, Automated Pharmacy Record Linkage in The Netherlands, Tayside Medicines Monitoring Unit (MEMO), and UK General Practice Research Database (GPRD). Most are databases of paid billing claims. The general advantages and disadvantages of these approaches were discussed in Chapter 11.

Over the past few decades, such databases have become a central means of performing hypothesis-testing studies in pharmacoepidemiology. In Chapter 13 we will briefly describe still other sources of pharmacoepidemiology data, and then in Chapter 14 we will place into perspective all the data sources described in Chapters 7–13, describing how to choose among them.

Key Points

Group Health Cooperative

- Data from health maintenance organizations (HMOs) have been used extensively to evaluate drug usage and the adverse and beneficial effects of marketed drugs and medical procedures.
- Group Health Cooperative (GHC), a nonprofit consumer-directed HMO established in 1947, currently provides health care on a prepaid basis to approximately 562 000 persons in Washington State.
- GHC's automated and manual databases serve as major resources for many epidemiologic studies, in part because individual records can be linked through time and across data sets by the unique consumer number assigned to each enrollee.
- Strengths of GHC as a setting for epidemiological research include its identifiable and relatively stable population base, accessible and complete medical records for each enrollee, and computerized databases.
- GHC databases still may be too small to detect associations between drug exposures and rare outcomes, necessitating combining data from multiple health care delivery systems, as has been done in the case of the Centers for Education and Research on Therapeutics.
- The GHC formulary limits the study of many newly marketed drugs.

Kaiser Permanente Medical Care Program

- Data from the Kaiser Permanente (KP) Medical Care Program provide information on over 8.2 million patients that are enrolled in one of the oldest and largest prepaid HMO in the US, covering eight states.
- Patient records across multiple databases (pharmacy records with hospitalizations, outpatient laboratory results, claims received from non-KP providers, etc.) and across time (for at least 10 years) can be linked, using a unique medical record number assigned to each patient for all encounters with the program.
- Cohort studies with considerable follow-up (and case-control studies with similar lengths of follow-back) are feasible in the KP database because of its size, diversity, representativeness, relative stability, and richness of its computerized clinical data.
- The KP clinical databases offer an advantage of allowing investigators to identify and adjust for many confounders.
- Researchers using this data resource must still be cognizant of certain limitations, such as: absence of complete, standard information on race/ethnicity or other indicators of socioeconomic status for all KP members; incomplete capture of all outpatient diagnoses; restrictive formularies; slower incorporation of some newer drugs compared with the fee-for-service environment; and reliance on records of prescriptions filled, which are not perfect measures of drug consumption.

HMO Research Network

- Research units in US HMOs are in a unique position to integrate research and practice for the improvement of health and health care among diverse populations.
- The combined study populations from HMOs provide sufficiently large sample size with a wide range of comorbid conditions and con-comitant medications to evaluate the beneficial and adverse effects of drugs.
- The HMO Research Network data source shares similar strengths and limitations with other claims-based systems in the US.

UnitedHealth Group

- UnitedHealth Group databases provide extensive data sources to study postmarketing drug safety and to evaluate risk communication efforts.
- Geographically diverse health plans facilitate pharmacoepidemiologic evaluation of the use of prescription drugs in general medical practice.
- The longitudinal nature of the data, ability to link sites of care, and large database size facilitate the study of rare exposures and rare outcomes.
- These databases have been used extensively to study drug safety and adverse drug reactions, augmented with medical record abstraction as applicable for study design.

Medicaid Databases

- The US Medicaid program provides medical coverage for certain categories of disadvantaged individuals.
- Data on prescription drugs are audited to detect fraud, and have been found to be accurate.
- The accuracy of diagnostic codes depends on the specific condition. With few exceptions, researchers using Medicaid data should verify diagnoses using primary medical records.

Health Services Databases in Saskatchewan

- Subjects are identified by a unique health services number that is used in each of the health care services databases (e.g., prescription drug data, hospitalizations, physician services) and can be used to link records across the databases and longitudinally.
- Data are population-based and cover most of the province's one million residents.
- Data have a long tenure (over 30 years) which enables compilation of extensive information about drug use and disease experience.
- Medical records in hospitals are accessible following established protocols, with availability ranging from 95% to 100% depending on the age of the record.
- Databases are well-suited to many exposure–outcome studies but limitations must be considered (e.g., population may be too small to study rare events; lack of data on some important confounders).

Automated Pharmacy Record Linkage in the Netherlands

- There have been strong regulatory and reimbursement incentives for Dutch patients to frequent a single GP and pharmacy, enabling the compilation of individual prescription drug and medical histories.
- The Dutch health care landscape can be considered due to its size, organization, information technology sophistication, and linkage to hospital and laboratory data as a “population laboratory”.
- Typical Dutch pharmacoepidemiology highlights are exposure characterization, channeling bias, and exposure time-windows.

Tayside Medicines Monitoring Unit (MEMO)

- The MEMO database is a community-based database reflecting a “real-world” population and representing all socioeconomic groups.
- The MEMO database is a validated record-linkage database.
- The core of the MEMO resource is a unique database that contains longitudinal data of dispensed prescribing and health outcomes over a 15-year period.
- MEMO has the capabilities to do a wide variety of population-based research projects such as pharmacoepidemiology, pharmacoconomics research, pharmacogenetics, and outcome research.

UK General Practice Research Database

- The General Practice Research Database (GPRD) contains anonymized data on diagnoses, therapies, and health-related behaviors recorded by GPs as part of the patients' electronic medical record.
- The GPRD is broadly representative of the UK and has population-based data on over 9 million patients, with over 44 million years of follow-up time allowing researchers to investigate rare outcomes.
- The GPRD has been validated and used to study a variety of medical conditions including cardiovascular, cancer, intestinal, dermatologic, pulmonary, and ophthalmologic outcomes.
- The GPRD may have incomplete information on some data from specialists as well as health-related behaviors. Investigators may obtain additional information by sending questionnaires to GPs through third-party vendors.
- The size and complexity of the GPRD requires that individuals or institutions working with it have adequate computer software and hardware, as well as experienced data managers. Alternatively, the GPRD can be accessed through a web version at www.gprd.com.

SUGGESTED FURTHER READINGS

GROUP HEALTH COOPERATIVE

- Barlow W, Davis RL, Glasser J, 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.
- Boudreau D, Leveille S, Gray S, Black DJ, Guralnik JM, Ferrucci L *et al.* Risks for frequent antimicrobial-treated infections in postmenopausal women. *Aging Clin Exp Res* 2003; **15**: 12–18.
- Clark DO, Von Korff M, Saunders K, Baluch B, Simon G. A chronic disease score with empirically derived weights. *Med Care* 1995; **33**: 783–95.
- Davis R, Kramarz P, Bohlke K, Benson P, Thompson RS, Mullooly J *et al.* Measles–mumps–rubella and other measles-containing vaccines do not increase the risk for inflammatory bowel disease: a case–control study from the Vaccine Safety Datalink project. *Arch Pediatr Adolesc Med* 2001; **155**: 354–9.
- Donahue J, Fuhlbrigge A, Finkelstein J, Fagan J, Livingston JM, Lozano P *et al.* Asthma pharmacotherapy and utilization by children in 3 managed care organizations. The Pediatric Asthma Care Patient Outcomes Research Team. *J Allergy Clin Immunol* 2000; **106**: 1108–14.
- Fishman P, Wagner EH. Managed care data and public health: the experience of Group Health Cooperative of Puget Sound. *Annu Rev Public Health* 1998; **19**: 477–91.
- Fishman P, Goodman M, Hornbrook M, Meenan RT, Bachman DJ, O’Keefe-Rosetti MC. Risk adjustment using automated ambulatory pharmacy data: the RxRisk model. *Med Care* 2003; **41**: 84–99.
- Goodwin F, Fireman B, Simon G, Hunkeler E, Lee J, Revicki D. Suicide risk in bipolar disorder during treatment with lithium and divalproex. *JAMA* 2003; **290**: 1467–73.
- Harris B, Stergachis A, Ried LD. The effect of drug copayments on the use and cost of pharmaceuticals in a health maintenance organization. *Med Care* 1990; **28**: 907–17.
- Jackson L, Neuzil K, Yu O, Benson P, Barlow WE, Adams AL *et al.* Effectiveness of pneumococcal polysaccharide vaccine in older adults. *N Engl J Med* 2003; **348**: 1747–55.
- Newton K, Wagner E, Ramsey S, 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.
- Pearson DC, Grothaus L, Thompson RS, Wagner EH. Smokers and drinkers in a health maintenance organization population: lifestyles and health status. *Prev Med* 1987; **16**: 783–95.
- 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.
- 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.

- Psaty BM, Koepsell TD, Siscovick D, Wahl P, Wagner EH. An approach to several problems in the use of large databases for population-based case–control studies of the therapeutic efficacy and safety of anti-hypertensive medicines. *Stat Med* 1991; **10**: 653–62.
- Psaty BM, Heckbert SR, Koepsell TD, Siscovick DS, Raghunathan TE, Weiss NS *et al.* The risk of myocardial infarction associated with anti-hypertensive drug therapies. *JAMA* 1995; **274**: 620–5.
- Rutter C, Mandelson M, Laya M, Seger D, Taplin S. Changes in breast density associated with initiation, discontinuation, and continuing use of hormone replacement therapy. *JAMA* 2001; **285**: 171–6.
- Simon G, Cunningham M, Davis R. Outcomes of prenatal antidepressant exposure. *Am J Psychiatry* 2002; **159**: 2055–61.
- Stergachis A, Shy K, Grothaus L, Wagner EH, Hecht JA, Anderson G *et al.* Tubal sterilization and the long-term risk of hysterectomy. *JAMA* 1990; **264**: 2893–8.
- 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.

KAISER PERMANENTE MEDICAL CARE PROGRAM

- 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.
- Collen MF, Davis LF. The multitest laboratory in health care. *J Occup Med* 1969; **11**: 355–60.
- Ferrara A, Quesenberry CP, Karter AJ, Njoroge CW, Jacobs AS, Selby JV. 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.
- Friedman GD, Collen MF, Harris LE, Van Brunt EE, Davis LS. Experience in monitoring drug reactions in outpatients: the Kaiser-Permanente Drug Monitoring System. *JAMA* 1971; **217**: 567–72.
- 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.
- Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *Am J Public Health* 1992; **82**: 703–10.
- Phipps KR, Stevens VJ. Relative contribution of caries and periodontal disease in adult tooth loss for an HMO dental population. *J Public Health Dent* 1995; **55**: 250–2.
- Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**: 915–20.

- Selby JV, Friedman GD. Screening prescription drugs for possible carcinogenicity: eleven to fifteen years of follow-up. *Cancer Res* 1989; **49**: 5736–47.
- Selby JV, Karter AJ, Ackerson LM, Ferrara A, Liu J. Developing a prediction rule from automated clinical data bases to identify high risk patients in a large population with diabetes. *Diabetes Care* 2001; **24**: 1547–55.
- Sidney S, Petitti DB, Quesenberry CP Jr. Myocardial infarction and the use of estrogen and estrogen/progestogen therapy in postmenopausal women. *Ann Intern Med* 1997; **227**: 501–8.
- 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.
- Van Den Eeden SK, Friedman GD. Prescription drug screening for subsequent carcinogenicity. *Pharmacoepidemiol Drug Saf* 1995; **4**: 275–87.
- Vogt TM, Elston Lafata J, Tolsma D, Greene SM. The role of research in integrated health care systems: the HMO Research Network. *Am J Manag Care* 2004; **10**: 643–8.

HMO RESEARCH NETWORK

- Califf R. The Centers for Education and Research on Therapeutics. The need for a national infrastructure to improve the rational use of therapeutics. *Pharmacoepidemiol Drug Saf* 2002; **11**: 319–27.
- 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 Org* 2000; **78**: 186–94.
- Krieger N, Chen JT, Waterman PD, Soobader M-J, Subramanian SV, Carson R. Geocoding and monitoring US socioeconomic inequalities in mortality and cancer incidence: does choice of area-based measure and geographic level matter?—the Public Health Disparities Geocoding Project. *Am J Epidemiol* 2002; **156**: 471–82.
- 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 in Therapeutics. *Pharmacoepidemiol Drug Saf* 2001; **10**: 373–7.
- Ray WA, Maclure M, Guess HA, Rothman KJ. Inception cohorts in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2001; (Suppl 1): 64–5.
- Selby J, Fraser I, Gunter M, Peterson E, Martinson B. *Results from IDSRN rapid cycle research Projects*. Abstract presented at the 9th Annual HMO Research Network Conference, April 2, 2003. Available from http://www.hmoresearchnetwork.org/archives/2003abst/03_ca_a4.pdf (accessed Dec 5, 2004).
- 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.
- Vogt TM, Elston-Lafata J, Tolsma D, Greene SM. The role of research in integrated healthcare systems: the HMO Research Network. *Am J Manag Care* 2004; **10**: 643–8.
- Wagner AK, Chan KA, Dashevsky I, Raebel MA, Andrade SE, Elston Lafata J *et al*. FDA drug prescribing warnings: is the black box half empty or half full? *Pharmacoepidemiol Drug Saf* 2006 (in press) DOI: 10.1002/pds.1193.
- 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; **35**: 3–11.

UNITEDHEALTH GROUP

- Andrade SE, Graham DJ, Staffa JA, Schech SD, Shatin D, La Grenade L, Goodman MJ, Platt R, Gurwitz JH, Chan KA. Health plan administrative databases can efficiently identify serious myopathy and rhabdomyolysis. *J Clin Epidemiol* 2005; **58**: 171–4.
- Gardner JS, Blough D, Drinkard CR, Shatin D, Anderson G, Graham D, Alderfer R. Tramadol and seizures: a surveillance study in a managed care population. *Pharmacotherapy* 2000; **20**: 1423–31.
- Graham DJ, Drinkard CR, Shatin D, Tsong Y, Burgess MJ. Liver enzyme monitoring in patients treated with troglitazone. *JAMA* 2001; **286**: 831–3.
- Graham DJ, Drinkard CR, Shatin D. Incidence of idiopathic acute liver failure and hospitalized liver injury in patients treated with troglitazone. *Am J Gastroenterol* 2003; **98**: 175–9.
- Graham DJ, Staffa JA, Shatin D, Andrade SE, Schech SD, La Grenade L, Gurwitz JH, Chan KA, Goodman MJ, Platt R. Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs. *JAMA* 2004; **292**: 2585–90.
- 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.
- Levy DG, Stergachis A, McFarland LV, Van Vorst K, Graham DJ, Johnson ES, Park BJ, Shatin D, Clouse JC, Elmer GW. Antibiotics and *Clostridium difficile* diarrhea in the ambulatory care setting. *Clin Ther* 2000; **22**: 91–102.
- McCarthy DB, Shatin D, Drinkard CR, Kleinman JH, Gardner JS. Medical records and privacy: empirical effects of legislation. *Health Serv Res* 1999; **34**: 417–25.
- Quam L, Ellis LBM, Venus P, Clouse J, Taylor CG, Leatherman S. Using claims data for epidemiologic research: the concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. *Med Care* 1993; **31**: 498–507.
- Rawson NSB, Shatin D. Assessing the validity of diagnostic data in large administrative health care utilization databases. In:

- Hartzema AG, Tilson HH, Chan KA, eds. *Pharmacoepidemiology and Therapeutic Risk Management*, 4th edn. Cincinnati: Harvey Whitney (in press).
- Rawson NSB, Nourjah P, Grosser SC, Graham DJ. Factors associated with celecoxib and rofecoxib utilization. *Ann Pharmacother* 2005; **39**: 597–602.
- Rector TS, Wickstrom SL, Shah N, Greenlee NT, Rheault P, Rogowski J, Freedman V, Adams J, Escarce JJ. Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions. *Health Service Res* 2004; **39**: 1839–57.
- Shatin D. Organizational context and taxonomy of health care databases. *Pharmacoepidemiol Drug Saf* 2001; **10**: 367–71.
- Shatin D, Drinkard CR. Ambulatory use of psychotropics by employer-insured children and adolescents in a national managed care organization. *Ambul Pediatr* 2002; **2**: 111–19.
- Shatin D, Gardner JS, Stergachis A, Blough D, Graham D. Impact of mailed warning to prescribers on the co-prescription of Tramadol and antidepressants. *Pharmacoepidemiol Drug Saf* 2005; **14**: 149–54.
- Shatin D, Rawson NSB, Stergachis A. UnitedHealth Group. In: Strom BL, ed. *Pharmacoepidemiology*. 4th edn. Chichester: Wiley, 2005; pp. 271–80.
- Shatin D, Schech SD, Brinker A. Ambulatory use of ticlopidine and clopidogrel in association with percutaneous coronary revascularization procedures in a national managed care organization. *J Interv Cardiol* 2002; **15**: 181–6.
- Shatin D, Rawson NSB, Curtis JR, Braun MM, Martin CK, Moreland LW, Becker AF, Patkar NM, Allison JJ, Saag KG. Documented tuberculin skin testing among infliximab users following multi-modal risk communication interventions. *Pharmacoepidemiol Drug Saf* 2006; **15**: 11–18.
- Willy ME, Manda B, Shatin D, Drinkard C, Graham D. A study of compliance with FDA recommendations for Pemoline (Cylert®). *J Am Acad Child Adolesc Psychiatry* 2002; **41**: 785–90.

MEDICAID DATABASES

- 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**: 1070.
- 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.
- National Pharmaceutical Council. *Pharmaceutical Benefits under State Medical Assistance Programs*. Reston, VA: National Pharmaceutical Council, 2003.
- 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.
- US Department of Health and Human Services. *A Profile of Medicaid*. Washington, DC: US Department of Health and Human Services, 2000.

HEALTH SERVICES DATABASES IN SASKATCHEWAN

- Albright PS, Livingstone S, Keegan DL, Ingham M, Shrikhande S, LeLorier J. Reduction in healthcare resource utilisation and costs following the use of risperidone for patients with schizophrenia previously treated with standard antipsychotic therapy. A retrospective analysis using the Saskatchewan Health linkable databases. *Clin Drug Invest* 1996; **11**: 289–99.
- Beck P, Wysowski D, Downey W, Butler-Jones D. Statin use and the risk of breast cancer. *J Clin Epidemiol* 2003; **56**: 280–5.
- Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of β_2 -agonists. *Am J Epidemiol* 1996; **144**: 1161–9.
- Bourgault C, Rainville B, Suissa S. Antihypertensive drug therapy in Saskatchewan: patterns of use and determinants in hypertension. *Arch Intern Med* 2001; **161**: 1873–9.
- Caro JJ, Migliaccio-Walle K, for CAPRA. Generalizing the results of clinical trials to actual practice: the example of clopidogrel therapy for the prevention of vascular events. *Am J Med* 1999; **107**: 568–72.
- Caro JJ, Salas M, Speckman JL, Raggio G, Jackson JD. Persistence with treatment for hypertension in actual practice. *Can Med Assoc J* 1999; **160**: 31–7.
- Collet JP, Sharpe C, Belzile E, Boivin JF, Hanley J, Abenhaim L. Colorectal cancer prevention by non-steroidal anti-inflammatory drugs: effects of dosage and timing. *Br J Cancer* 1999; **81**: 62–8.
- Csizmadia I, Collet J-P, Benedetti A, Boivin J-F, Hanley JA. The effects of transdermal and oral oestrogen replacement therapy on colorectal cancer risk in postmenopausal women. *Br J Cancer* 2004; **90**: 76–81.
- Hemmelgarn B, Blais L, Collet, JP, Ernst P, Suissa S. Automated databases and the need for fieldwork in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 1994; **3**: 275–82.
- Joffe RT, Iskredjian M, Einarson TR, O'Brien BJ, Stang MR. Examining the Saskatchewan health drug database for antidepressant use: the case of fluoxetine. *Can J Clin Pharmacol* 2001; **8**: 146–52.
- Johnson J, Majumdar S, Simpson S, Toth E. Decreased mortality associated with the use of metformin compared with sulfonylurea monotherapy in Type 2 diabetes. *Diabetes Care* 2002; **25**: 2244–8.
- Liu L, Reeder B, Shuaib A, Mazagri R. Validity of stroke diagnosis on hospital discharge records in Saskatchewan, Canada: implications for stroke surveillance. *Cerebrovasc Dis* 1999; **9**: 224–30.
- Rawson NSB, Malcolm E. Validity of the recording of ischaemic heart disease and chronic obstructive pulmonary disease in the Saskatchewan health care datafiles. *Stat Med* 1995; **14**: 2627–43.
- Ray WA, Griffin MR, Downey W, Melton LJ. Long-term use of thiazide diuretics and risk of hip fracture. *Lancet* 1989; **i**: 687–90.
- Sharpe CR, Collet JP, McNutt M, Belzile E, Boivin JF, Hanley JA. Nested case-control study of the effects of non-steroidal anti-inflammatory drugs on breast cancer and stage. *Br J Cancer* 2000; **83**: 112–20.

- Simpson S, Corabian P, Jacobs P, Johnson J. The cost of major comorbidity in people with diabetes mellitus. *Can Med Assoc J* 2003; **168**: 1661–7.
- Spitzer WO, Suissa S, Ernst P, Horwitz RI, Habbick B, Cockcroft D *et al.* The use of β -agonists and the risk of death and near death from asthma. *N Engl J Med* 1992; **326**: 501–6.
- Stang MR, Wysowski DK, Butler-Jones D. Incidence of lactic acidosis in metformin users. *Diabetes Care* 1999; **22**: 925–7.
- Tennis P, Bombardier C, Malcolm E, Downey W. Validity of rheumatoid arthritis diagnoses listed in the Saskatchewan hospital separations database. *J Clin Epidemiol* 1993; **46**: 675–83.
- Wang EEL, Einarson TR, Kellner JD, Conly JM. Antibiotic prescribing for Canadian preschool children: evidence of over-prescribing for viral respiratory infections. *Clin Infect Dis* 1999; **29**: 155–60.

AUTOMATED PHARMACY RECORD LINKAGE IN THE NETHERLANDS

- Beiderbeck-Noll AB, Sturkenboom MC, van der Linden PD, Herings RM, Hofman A, Coebergh JW *et al.* Verapamil is associated with an increased risk of cancer in the elderly: the Rotterdam study. *Eur J Cancer* 2003; **39**: 98–105.
- De Bruin ML, van Puijenbroek EP, Egberts AC, Hoes AW, Leufkens HG. Non-sedating antihistamine drugs and cardiac arrhythmias—biased risk estimates from spontaneous reporting systems? *Br J Clin Pharmacol* 2002; **53**: 370–4.
- Egberts ACG, Lenderink AW, Koning de FHP, Leufkens HGM. Channeling of three newly introduced antidepressants to patients not responding satisfactorily of previous treatment. *J Clin Psychopharmacol* 1997; **17**: 149–55.
- Heerdink ER, Leufkens HG, Herings RM, Ottervanger JP, Stricker BH, Bakker A. NSAIDs associated with increased risk of congestive heart failure in elderly patients taking diuretics. *Arch Intern Med* 1998; **158**: 1108–12.
- Herings RMC, Stricker BHC, Nap G, Bakker A. Pharmacomorbidity linkage: a feasibility study comparing morbidity in two pharmacy-based exposure cohorts. *J Epidemiol Community Health* 1992; **46**: 136–40.
- Herings RM, Urquhart J, Leufkens HG. Venous thromboembolism among new users of different oral contraceptives. *Lancet* 1999; **354**: 127–8.
- Hoes AW, Grobbee DE, Lubsen J, Man in't Veld AJ, van der Does E, Hofman A. Diuretics, beta-blockers, and the risk for sudden cardiac death in hypertensive patients. *Ann Intern Med* 1995; **123**: 481–7.
- Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* 1991; **7**: 403–22.
- Maitland-van der Zee AH, de Boer A, Leufkens HG. The interface between pharmacoepidemiology and pharmacogenetics. *Eur J Pharmacol* 2000; **410**: 121–30.
- Monster TB, Janssen WM, de Jong PE, de Jong-van den Berg LT. PREVEND Study Group. The impact of antihypertensive drug groups on urinary albumin excretion in a non-diabetic population. *Br J Clin Pharmacol* 2002; **53**: 31–6.
- Petri H, Leufkens H, Naus J, Silkens R, Hessen P van, Urquhart J. Rapid method for estimating the risk of acutely controversial side effects of prescription drugs. *J Clin Epidemiol* 1990; **43**: 433–9.
- Schoofs MW, van der Klift M, Hofman A, de Laet CE, Herings RM, Stijnen T *et al.* Thiazide diuretics and the risk for hip fracture. *Ann Intern Med* 2003; **139**: 476–82.
- Souverein PC, Egberts AC, Sturkenboom MC, Meuleman EJ, Leufkens HG, Urquhart J. The Dutch cohort of sildenafil users: baseline characteristics. *Br J Urology* 2001; **87**: 648–53.
- Straus SM, Bleumink GS, Dieleman JP, van der Lei J, 't Jong GW, Kingma JH *et al.* Antipsychotics and the risk of sudden cardiac death. *Arch Intern Med* 2004; **164**: 1293–7.
- Van der Lei J, Duisterhout JS, Westerhof HP *et al.* The introduction of computer-based patient records in The Netherlands. *Ann Intern Med* 1993; **119**: 1036–41.
- van Puijenbroek E, Diemont W, van Grootheest K. Application of quantitative signal detection in the Dutch spontaneous reporting system for adverse drug reactions. *Drug Saf* 2003; **26**: 293–301.

TAYSIDE MEDICINES MONITORING UNIT (MEMO)

- 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.
- Beardon PH, McGilchrist MM, McKendrick AD, McDevitt DG, MacDonald TM. Primary non-compliance with prescribed medication in primary care. *BMJ* 1993; **307**: 846–8.
- Carstairs V. Deprivation and health in Scotland. *Health Bull (Edinb)* 1990; **48**: 162–75.
- Stationary Office. *Data Protection Act*. London: The Stationary Office, 1998; Chapter 29.
- Doney AS, Fischer B, Cecil JE, Boylan K, McGuigan FE, Ralston SH *et al.* Association of the Pro12Ala and C1431T variants of PPARG and their haplotypes with susceptibility to Type 2 diabetes. *Diabetologia* 2004; **47**: 555–8.
- International Classification of Disease. *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death*, 9th revision, Vol. 1. Geneva: World Health Organization, 1977.
- International Society for Pharmacoepidemiology. *Guidelines for Good Pharmacoepidemiology Practices (GPP)*. Available at: http://www.pharmacoepi.org/resources/guidelines_08027.cfm/.
- Leese GP, Wang J, Broomhall J, Kelly P, Marsden A, Morrison W *et al.* frequency of severe hypoglycemia requiring emergency treatment in Type 1 and Type 2 diabetes: a population-based study of health service resource use. *Diabetes Care* 2003; **26**: 1176–80.
- Libby G, Smith A, McEwan NF, Chien PFW, Greene SA, Forsyth JS *et al.* The Walker Project: a longitudinal study of 48,000 children born 1952–66 (aged 36–50 years in 2002) and

- their families. Study methodology. *Paediatr Perinat Epidemiol* 2004; **18**: 302–12.
- MacDonald TM, Wei Li. Effect of ibuprofen on cardioprotective effect of aspirin. *Lancet* 2003; **361**: 573–4.
- MacDonald TM, Morant SV, Robinson GC, Shield MJ, McGilchrist MM, Murray FE *et al*. Association of upper gastrointestinal toxicity of non-steroidal anti-inflammatory drugs with continued exposure: cohort study. *BMJ* 1997; **315**: 1333–7.
- MacDonald TM, Morant SV, Goldstein JL, Burke TA, Pettitt D. Channelling bias and the incidence of gastrointestinal haemorrhage in users of meloxicam, coxibs, and older, non-specific nonsteroidal anti-inflammatory drugs. *Gut* 2003; **52**: 1–6.
- 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.
- Office of Population Censuses and Surveys. *Tabular List of the Classification of Surgical Operations and Procedures*, 4th revision. London: HMSO, 1990.
- Scottish Health Statistics. Appendixes. Available at: http://www.show.scot.nhs.uk/isdonline/Scottish_Health_Statistics/SHS-98/Appendix.pdf/.
- Sheen CL, Dillon JF, Bateman DN, Simpson KJ, Macdonald TM. Paracetamol toxicity: epidemiology, prevention and costs to the health-care system. *Q J Med* 2002; **95**: 609–19.
- Wang J, Donnan PT, Steinke D, MacDonald TM. The multiple propensity score for analysis of dose–response relationships in drug safety studies. *Pharmacoepidemiol Drug Saf* 2001; **10**: 105–11.
- Wei L, MacDonald TM, Davey PG. Relation between socioeconomic deprivation and statin prescribing in post myocardial infarction patients: a population based study. *Pharmacoepidemiol Drug Saf* 2002; **11** (Suppl): S151.
- Wei L, MacDonald TM, Walker BR. Taking glucocorticoids by prescription is associated with subsequent cardiovascular disease. *Ann Intern Med* 2004; **141**: 764–70.
- Wei L, Ebrahim S, Bartlett C, Davey PG, Sullivan FM, MacDonald TM. Cohort study of statin use in the secondary prevention of coronary heart disease in primary care: comparison of inclusion and outcome with trial data. *BMJ* 2005; **330**: 821–4.
- GENERAL PRACTICE RESEARCH DATABASE
- Gelfand JM, Berlin J, Van Voorhees A, Margolis DJ. Lymphoma rates are low but increased in patients with psoriasis: results from a population-based cohort study in the United Kingdom. *Arch Dermatol* 2003; **139**: 1425–9.
- Gelfand JM, Wang X, Liu Q, Neimann AN, Weinstein R, Margolis DJ, Troxel AB. Epidemiology and treatment patterns of psoriasis in the General Practice Research Database (GPRD). *Pharmacoepidemiol Drug Saf* 2005; **14**: S23.
- Hollowell J. The General Practice Research Database: quality of morbidity data. *Popul Trends* 1997; **87**: 36–40.
- Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ* 1991; **302**: 766–8.
- Jick SS, Kaye JA, Vasilakis-Scaramozza C, Garcia Rodriguez LA, Ruigomez A, Meier CR, Schlienger RG, Black C, Jick H. Validity of the general practice research database. *Pharmacotherapy* 2003; **23**: 686–9.
- Lawson DH, Sherman V, Hollowell J. The General Practice Research Database. Scientific and Ethical Advisory Group. *Q J Med* 1998; **91**: 445–52.
- Lewis JD, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. *Pharmacoepidemiol Drug Saf* 2004; **13**: 437–41.
- Margolis DJ, Knauss J, Bilker W. Hormone replacement therapy and prevention of pressure ulcers and venous leg ulcers. *Lancet* 2002; **359**: 675–7.
- Margolis DJ, Bowe WP, Hoffstad O, and Berlin JA. Antibiotic treatment of acne may be associated with upper respiratory tract infections. *Arch Dermatol* 2005; **141**: 1132–6.
- Metlay JP, Kinman JL. Failure to validate pneumococcal pneumonia diagnoses in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2003; **12**: S163.
- Walley T, Mantgani A. The UK General Practice Research Database. *Lancet* 1997; **350**: 1097–9.
- Wood L, Coulson R. Revitalizing the General Practice Research Database: plans, challenges, and opportunities. *Pharmacoepidemiol Drug Saf* 2001; **10**: 379–83.

13

Other Approaches to Pharmacoepidemiology Studies

Edited by:

BRIAN L. STROM

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

INTRODUCTION

As described in Chapter 3, although pharmacoepidemiology studies use the traditional study designs of epidemiology, they pose a special problem. Inasmuch as most drugs are studied in between 500 and 3000 individuals prior to marketing, postmarketing surveillance cohort studies will not be able to detect rarer drug effects reliably unless they include at least 10 000 exposed individuals. Postmarketing surveillance case-control studies need to accumulate diseased cases and undiseased controls from a target population of sufficient size to have included 10 000 exposed patients if a cohort study had been performed. The need to study populations this large without incurring undue costs represents an unusual logistical challenge.

Some of the major approaches useful for addressing this challenge have been presented in Chapters 7–12. There are a number of other approaches as well. Several of these are derived from one or more of the approaches presented earlier, but some represent important potential resources that the reader should be aware of. Thus, the purpose of this chapter is to describe them. The approaches will

be presented according to the study designs they usually utilize, in order of the hierarchy of study designs presented in Chapter 2. Approaches will be presented that involve performing analyses of secular trends, case-control studies, cohort studies, and randomized clinical trials. None of these approaches involves analyzing case reports. Case series will be discussed under case-control and cohort studies, since exposed (or diseased) patients in case series should generally be compared to unexposed (or undiseased) controls.

DATA SOURCES FOR ANALYSES OF SECULAR TRENDS

As described in Chapter 2, analyses of secular trends examine trends in an exposure and trends in a disease and explore whether the trends coincide. These trends can be examined over time or over geographic area. In other words, one could analyze data from a single country or region and examine how the exposure and the disease have changed over time. Alternatively, one could analyze data from a single time

period, exploring how the prevalence of the exposure and the incidence of the disease differ from region to region or country to country.

The advantages and disadvantages of this study design are presented in Chapter 2. Analyses of secular trends in pharmacoepidemiology are *ad hoc* studies, i.e., there are no ongoing systems for performing such studies. In order to perform such studies, however, one obviously needs data on both the frequency of the exposure and the incidence of the disease. Thus, this discussion will focus on the data sources available for performing such studies in pharmacoepidemiology.

DRUG UTILIZATION DATA

In pharmacoepidemiology studies, the primary exposure of interest is drug use. Thus, in order to perform analyses of secular trends, one needs to have one or more sources of data on drug utilization. In the US, and in many other countries, the major sources of data on drug utilization are several private companies that specialize in collecting these data and then selling them to pharmaceutical manufacturers for use in marketing studies.

Probably the best-known source of such data is IMS HEALTH, a leading data resource on the sales of pharmaceuticals worldwide. IMS conducts a number of different ongoing surveys of drug use, which it then sells to pharmaceutical manufacturers and to government agencies, as well as other categories of clients. Perhaps the most useful of these for pharmacoepidemiologic research is the National Disease and Therapeutic Index™ (NDTI™), an ongoing medical audit that provides insight into disease and treatment patterns of office-based physicians. A rotating panel of over 3500 of the approximately 400 000 office-based US physicians reports four times each year on all contacts with patients during a 48-hour period. Data are collected on the drug prescribed and its quantity, the diagnosis the drug was prescribed for, the action desired, concomitant drugs, concomitant diagnoses, and whether the prescription in question was the first time the patient received the drug or whether it was for continuing therapy. Demographic data about the patient and the prescriber are also collected. Periodic reports are prepared, including reports organized by drug and by diagnosis. In addition, special analyses can be performed using the data.

Although the physician panel is relatively small compared to the overall total of office-based physicians in the US, NDTI™ data are projected to the national level using statistical methodology. The sample of physicians used for the

NDTI™ is a thin one, i.e., it contains relatively few individuals, relative to the number of individuals being generalized to and the number of variables being studied. Nevertheless, the results obtained using the NDTI™ tend to be relatively stable over time, and agree fairly well with other sources of drug utilization information. NDTI™ has proven itself very useful, and it has been used often for pharmacoepidemiologic research.

A few of the other IMS databases are: (i) the National Prescription Audit Plus™ (NPA Plus™), which is a study of pharmacy sales from retailers, mail order, and long-term care facilities, based on a panel of computerized pharmacies which submit data on dispensed prescriptions to IMS electronically; (ii) LifeLink™, which is a patient longitudinal database of medical and pharmacy claims; and (iii) IMS National Sales Perspective™—Retail and National Sales Perspective™—Non-Retail, two audits that provide national sales dollar estimates of pharmaceutical products purchased by retail drug stores, food stores, mass merchandise, closed wall health maintenance organizations (HMOs), non-Federal hospitals, Federal facilities, clinics and long-term care facilities. The IMS National Sales Perspective™, based on data from both audits, covers over 90% of drug sales. The existence of these and other IMS databases changes over time, as new products are developed and old products are discontinued. These and other IMS databases can sometimes be useful for pharmacoepidemiologic research, but the NDTI™ is generally the most useful. Generously, IMS is often willing to make its data available to academic investigators at little or no cost.

There are other commercial sources of drug utilization data in the US, although these have not been used as frequently for academic research. In addition, as part of its National Ambulatory Medical Care Survey, the US National Center for Health Statistics has been studying drug use, providing data very similar to that included in IMS America's NDTI™. A summary of these results used to be published in an annual publication entitled *Highlights of Drug Utilization in Office Practice*, and other studies have been published from them since. Other special analyses can be requested as well. Other potential sources of drug utilization data include all of the databases described in Chapters 11 and 12.

A new source of drug utilization data is the Slone Survey. This is based on an ongoing telephone random survey of the non-institutionalized population in the continental US. Its first report contained data on 3180 participants interviewed in 1998 and 1999. It is an excellent new source of information about the US use of prescription drugs, but also non-prescription drugs, vitamins/minerals, and

herbals/supplements. Its limitations, of course, are nonparticipation, although the participation rate was an excellent 72%, and the uncertain and variable validity of the data obtained via a telephone interview with selective memory prompts (see Chapter 15).

Finally, in a number of countries other than the US, true national data are available on drug sales. For example, Sweden's Apoteksbolaget (now called Apotek AB), the National Corporation of Swedish Pharmacies, provides pharmacy services for the entire country. It retains remarkable data on the drugs dispensed in Sweden, regionally or nationally. Through analyses of this type of data, Scandinavian investigators have become the world's leaders in studies of drug utilization (see Chapter 27).

It is important to note that, with the exception of data derived from the databases described earlier in the book, all of the drug utilization databases are useful for descriptive studies only, not for analytic studies. In some drug utilization data sets, information on subsequent diagnoses is unavailable. For example, in NDTI™ the patients who happen to be seen in any given 48-hour cycle are unlikely to be the same as those in the preceding or subsequent 48-hour cycles. In some drug utilization data sets information *is* available on diagnoses, but may be inappropriate. For example, the diagnosis data in NDTI™ includes the indication for treatment rather than the results of treatment. Finally, in some drug utilization data sets the information on drugs cannot be linked to disease outcome. For example, the excellent Swedish drug data cannot be linked to the excellent Swedish hospitalization data, as the former are retained in aggregate only, that is, without individual patient identification numbers. Thus, this type of data can only be used for descriptive studies or aggregate analyses, such as those performed for analyses of secular trends. More information on drug utilization studies is provided in Chapter 27.

DISEASE INCIDENCE DATA

Data Sources

The major source of disease incidence information useful for this type of study are the vital statistics maintained by most countries in the world. Most countries, for example, maintain mortality statistics, derived from the death certificates completed by physicians at the time of death. Importantly, these death certificates include information on causes of death. In the US, these death certificates are collected and maintained by the states. However, the National Center for Health Statistics then obtains magnetic tapes of a portion of these data from the states' vital statistics offices. These

have been compiled into the National Death Index, which can be used for research purposes at a modest cost. Data are available beginning in 1979. The index contains the following information for each of those who died: last name, first name, middle initial, social security number, date of birth, state of birth, father's surname, sex, race, marital status, and state of residence. It also contains the name of the state, the death certificate number, and the date of death. To obtain cause of death information, an investigator must request a death certificate directly from the state. A similar data resource in Canada is the Canadian Mortality Database.

US death data are also available from the Social Security Administration, as the "Social Security death tapes." These arise out of the government's need to know of deaths, in order to avoid paying social security benefits to individuals who have died.

Other types of vital statistics that are recorded by most countries include birth data, marriage data, divorce data, and so on. These are less likely to be useful for pharmacoepidemiologic research.

Morbidity data can be more problematic. There is no comprehensive source of morbidity data in most countries, comparable to mortality data. However, many specific types of data are available. A large number of countries maintain cancer registries, collecting all cases of cancer in one or more defined populations. These can be used to calculate incidence rates of cancer. Other specialized registries can also exist. For example, in the US the Centers for Disease Control maintains a registry of children born with birth defects in the Atlanta metropolitan area (see Chapter 27).

In addition, many developed countries maintain "population laboratories," defined and stable populations which have been observed over time, with multiple measurements made of both exposures and diseases. The classic population laboratory in the US has been Framingham, Massachusetts, which has been the source of an enormous amount of knowledge about risk factors for cardiovascular disease, as well as other diseases.

Many developed countries also conduct periodic health surveys, in order to explore trends in exposures and diseases. For example, the US National Center for Health Statistics conducts a periodic Health Interview Study as a study of illnesses reported by patients. The National Center for Health Statistics also conducts a periodic Health Examination Survey, including physical examinations and laboratory tests. The National Ambulatory Medical Care Survey, mentioned above, investigates samples of office-based physicians. The Health Records Survey (no longer being conducted) and the Institutional Population Survey (no longer being conducted) investigated samples

of institutionalized patients. The Hospital Discharge Survey investigates samples of patients discharged from acute care hospitals. As a last example, the National Natality and Mortality Surveys collect additional data on individuals sampled from vital statistics data.

Finally, many countries have selected “reportable diseases.” These are diseases of particular interest to the local public health authorities, and are often infectious diseases. Sometimes reporting is “required,” although enforcement is difficult. Sometimes reporting is just requested. In either case, however, reporting is not complete, and so it can be difficult to disentangle trends in disease incidence from trends in reporting.

Potential Problems

This type of data can be extremely useful in performing analyses of secular trends, and has been used to address a number of important questions in pharmacoepidemiology. However, whenever one uses data that were not collected specifically for the study at hand, one must be very careful to be aware of its limitations. Each of the data sources described presents its own problems. Mortality data are the most likely to be used for analyses of secular trends, and the problems of these data represent good illustrations of the types of problems one must consider when using any of the data sources. As such, they will be discussed in more detail.

Overall, in using mortality data one is limited by the care, or lack of care, taken by physicians in completing death certificates. This is particularly a problem for studies that rely on information about the cause of death. Physicians may not know the cause of death accurately or, even if they do, they may not be careful in recording it. This is unlikely to create false findings in analyses of secular trends, unless there is a systematic change in these errors over time or across geographic areas. Unfortunately, however, these systematic changes can occur in many ways.

First, one can see changes in physicians’ index of suspicion about any given disease. This can lead to trends in how frequently patients are diagnosed with the disease. For example, pulmonary emboli frequently are not detected. As physicians have become aware of this problem, one would expect that a larger proportion of patients with this disease would be diagnosed.

Second, diagnostic methods can change over time. This, too, can create false trends. For example, clinical diagnoses of pulmonary emboli have been shown to be wrong over 50% of the time. As lung scanning procedures became widely available, one would expect that a larger proportion

of patients with pulmonary embolism received correct diagnoses.

Third, diagnostic terminology can change over time. For example, with the development of the extractable nuclear antigen serologic test, patients previously diagnosed with other conditions are now being diagnosed as having mixed connective tissue disease.

Fourth, there can be changes in coding systems. A blatant example is the periodic shift from an older version of the International Classification of Diseases to a more recent one, such as from ICD-8 to ICD-9. Less obvious changes can cause major problems as well. For example, in a study of methylodopa and biliary carcinoma using data from multiple international cancer registries, all but one showed no association between drug sales and disease incidence or mortality. In one, however, a marked increase shortly after drug marketing was seen in cancer of the biliary tract, excluding the code for cancer of the biliary tract—part unspecified. A complementary decrease was seen for the excluded code. Further investigation revealed that a change in coding policy was instituted in that registry in 1966. After that date, more specific coding was to be used. This resulted in an apparent increase in the incidence rates of the diseases of specified sites, accompanied by an apparent decrease in the incidence rates of diseases with unspecified sites.

Fifth, there can be changes in population demographics. The aging of the US population now underway would obviously dictate a shift in mortality from diseases of the young to diseases of the old. Age-specific analyses could be performed to control for this trend, but other trends cannot be adjusted for as easily, for example migration.

Finally, using mortality data one cannot differentiate between a change in the incidence rate of a disease and a change in the case-fatality rate of a disease. For example, we know that cardiovascular mortality is decreasing in much of the developed world. However, we do not know whether that decrease is because fewer people are developing coronary artery disease or whether the same proportion of the population is developing the disease but they are living longer with the disease before dying.

DATA SOURCES FOR CASE-CONTROL STUDIES

REGISTRY DATA

As discussed above, there are a number of registries available, each comprised of cases of selected diseases. Usually these are just a collection of cases, without controls.

However, if there is a registry extant that has a collection of cases of a disease one wishes to study, then this can be useful for performing a case-control study of that disease. One needs to be careful, however, about whether the registry collected all cases of a disease in a defined population or just some of them. If the latter, then one needs to consider whether the method of recruitment might introduce some bias into the study.

One particular registry, which is often forgotten in this context, is the spontaneous reporting system maintained by regulatory bodies throughout the world (see Chapters 7 and 8). These represent sources of cases of adverse drug reactions and can be used for case finding for case-control studies. As a specific example, a study used this approach to investigate the pathophysiology of the suprofen-induced flank pain syndrome. Cases with the acute flank pain syndrome were compared to controls without the syndrome, both groups exposed to suprofen. This provided interesting information on risk factors for the acute flank pain syndrome among those who have been exposed to suprofen.

OLMSTED COUNTY MEDICAL RECORDS

Another approach to performing case-control studies takes advantage of the unique medical record system in Olmsted County, Minnesota. The Mayo Clinic and its affiliated clinics and hospitals provide the medical care for most of the 125 000 residents of Olmsted County. These data have been supplemented since 1966 by information on diagnostic and surgical procedures from the other medical groups and hospitals in Olmsted County, as well as the few independent practitioners who were not part of the Mayo Clinic system. This database now covers the medical care delivered to County residents from 1909 through the present. These records represent an extremely useful and productive resource for epidemiologic studies, including case finding for case-control studies. For pharmacoepidemiology studies, however, drug exposure data must usually be gathered *de novo*.

MANITOBA

The Canadian province of Manitoba maintains computerized files for its clients, as a by-product of its provincial health plan. This administrative database has 1.1 million people, starting in 1970. Data can be accessed from four computerized databases maintained by Manitoba Health Services Insurance Plan (MHSIP). These linked databases include: (i) registration files, containing a record for every

individual registered to receive insured health services, (ii) records of physician reimbursement claims for medical care provided, containing information on patient diagnosis and physician specialty, (iii) records of hospital discharge abstracts, and (iv) records of prescriptions dispensed in retail pharmacies, containing data on the date of prescription dispensing, drug name, dosage form, quantity dispensed, and a drug identification number.

AD HOC CASE-CONTROL STUDIES

Finally, case-control studies can be performed as *ad hoc* studies as well. Cases can then be recruited from whatever source is appropriate for that disease, whether hospitals, outpatient practices, or some other source. Investigators in some geographic areas maintain ongoing relationships with a number of hospitals, to permit case finding for case-control studies.

Case definitions must, of course, be clear. Cases should be "incident cases," that is, individuals who have recently developed the disease, so one can inquire about exposures that precede the onset of the disease. All individuals who meet the case definition should be enrolled, if possible, to decrease the risk of a selection bias. Finally, if all cases can be identified in a defined population, then the incidence of the disease in the population can be determined.

Controls can then be recruited from either the site of medical care for the cases or from the community where the cases come from. The latter is now generally perceived as a better approach than the former. Community controls can be recruited from friends of the cases, from neighbors of the cases, from a reverse telephone directory, by using random digit dialing, or from some comprehensive listing of the target population. The last is generally the best approach, although it often is not available. Exceptions are selected geographic areas in which the government maintains such a list, organized medical care programs which can provide listings of eligible individuals, such as general practitioners' patient lists in the United Kingdom and the other programs described in Chapters 11 and 12, and other special situations. An example would be a study limited to the elderly, which could obtain listings of those eligible for Medicare in the local area.

Friend controls are convenient, but risk overmatching; friends may be similar in personal habits, and this can be problematic if these are risk factors of importance in the study. Reverse telephone directories list individuals by address, rather than by name. They can be useful for choosing community controls that are matched for neighborhood

and, thereby, crudely matched for socioeconomic status. However, the use of reverse telephone directories to recruit controls is problematic in areas where a large proportion of the population has unlisted telephone numbers. This is becoming common in major metropolitan areas in the US. Thus, random digit dialing is often the best method available for the selection of community controls. As shown by results from a recently published study, despite concerns about low participation rates in random digit dialing surveys, drug utilization information provided by study participants may be representative of the utilization practices of the nonparticipants. As part of a methodologic study designed to compare agreement between women's reports of their utilization of hormone replacement therapy (HRT) and the information available for these women in a claims database about drugs dispensed to them, selection bias was tested by assessing the difference in utilization of HRT between responders and nonresponders. Women aged 50–79 years old were contacted to ask them to participate in a telephone interview about their hormone use. An initial screening telephone call was administered in order to recruit them into the study and to arrange the time for the main telephone interview to be administered. The contact experience with the study subjects was designed to be similar to that for typical random digit dialing. Out of a random sample of 213 women selected from the claims database who were contacted, 154 (72.3%) women agreed to participate and 59 (32.7%) women refused. Among the 154 women who agreed to participate, 79 (51.3%, 95% CI: 43.1–59.4%) were shown by the claims database to have been dispensed an HRT during a 15-month period. Among the 59 women who refused to participate, 30 (50.8%, 95% CI: 37.5–64.1%) were shown by the database to have been dispensed an HRT during the same period. Thus, this study showed that use of HRT was almost identical in responders and nonresponders.

However, patient refusals are becoming an increasing problem for random digit dialing. While it is difficult to compare response rates among studies because of variations in calculating and reporting these rates, there is a sense in the field that recent rates are lower than rates reported in earlier studies. Further, the growing use of cell phones instead of landlines may make this problematic in the future as well, if the cell phone numbers remain inaccessible to researchers.

The validity of exposure data collected from patients as part of *ad hoc* case-control studies is discussed in Chapter 15.

Of course, many other details must be considered in planning a case-control study. A more extensive discussion of this is beyond the scope of this book. The interested reader

is referred to a standard epidemiology textbook and/or to one of the books now available which specifically discuss case-control studies. Overall, the advantages and disadvantages of this approach are identical to those of case-control surveillance, described in Chapter 9. The major additional advantages of the latter are that it has a large database of potential controls and a standardized procedure, which can expedite the study process. However, *ad hoc* studies have more flexibility in their design, allowing one to use community controls and to tailor the data collection effort to the question at hand.

DATA SOURCES FOR COHORT STUDIES

The major logistic issues in performing pharmacoepidemiology studies using a cohort design are, first, how to identify a cohort of patients exposed to a drug of interest and one or more control groups without exposure to the drug and, second, how to determine their clinical outcome. The major sources of information about drug exposures are billing claims, physicians, pharmacies, and patients. To date, the last has been considered relatively unreliable, and most approaches have used one of the other three sources. Examples of each will be presented below. As to clinical outcomes, the major source of information must be the physician, directly or indirectly, through the medical record. Patients can be used as a partial source of this information, but confirmatory and supplementary information will generally be needed from physicians. The validity of disease outcome data collected from different sources is discussed in more detail in Chapter 15.

PHARMACY-BASED POSTMARKETING SURVEILLANCE STUDIES

A relatively underused method of recruiting patients into pharmacoepidemiology studies is through the pharmacy that dispenses the drug. One approach to this would be to collect data on drug exposures from computerized pharmacies. This would be similar to a billing data source. Alternatively, one could obtain the participation of pharmacists, asking them to solicit patient recruitment. Finally, one could enclose recruitment information in a drug's packaging, requesting that patients return an enclosed business reply card to enroll in the study.

The major pioneer in the use of pharmacy-based methods of pharmacoepidemiology was the pharmacoepidemiology group at what used to be Upjohn Pharmaceuticals. The results of a feasibility study using this approach were

published. Briefly, the study consisted of the identification and follow-up of 21 372 patients treated between July 1975 and July 1977 with oral antibacterials in an ambulatory care setting. Participating centers were limited to those which combined medical care delivery sites with on-site pharmacies, excluding large hospital outpatient clinics and large referral centers. The pharmacists at participating sites were asked to invite a patient to participate if he or she received a drug of interest. The patient was given an explanatory brochure, which was supplemented by a discussion with the pharmacist, if needed. The brochure included a "Release of Information" statement, which was retained by the pharmacist. At weekly intervals, the pharmacist sent the coordinating center a list of all antibacterials dispensed, information about each prescription, and information about the patients who agreed to participate. The pharmacists were paid for the time this involved. Data on health outcomes were collected from the patients using a questionnaire mailed to them one month later. This was supplemented by telephone follow-up when needed. Reports of hospitalizations or deaths were confirmed at the place of treatment.

Pharmacy-based surveillance was used by Upjohn in other studies as well. In general, the approach remained the same, although computer-assisted telephone interviews were used to collect outcome data, rather than mailed questionnaires. People who could not be contacted by telephone were sent certified letters, asking them to telephone the research center. If they signed the receipt for the certified letter, but still could not be contacted, they were classified as alive.

The advantages of this approach are that it is free of the selection bias inherent in using physicians to recruit patients. Also, this approach does not interfere with prescribing practices: it allows one to collect information about patients' use of concomitant drugs other than those any given physician may be aware of, it allows one to study outcomes which need not come to medical attention, and, compared to studies that recruit patients via prescribers, it is less expensive—it is free of the large cost of reimbursing the prescriber for his cooperation. The disadvantages of the approach are the potential for a volunteer bias and the extensive resources and time needed for site recruitment and data collection. Overall, however, this appears to be a very effective, although underused, approach to performing cohort studies in pharmacoepidemiology.

As another approach to pharmacy-based surveillance, the Center for Medication Monitoring at the University of Texas Medical Branch in Galveston has been performing postmarketing surveillance using patient self-monitoring. Patients

filling a prescription for a target medication are presented with an announcement of the study along with their prescription. Patients who agree to participate are then asked to report during the next month on any changes in their health status. While they are mailed two questionnaires to obtain demographic and medication usage information, patients are asked to telephone the Center to report any new clinical events. This is a variation on pharmacy-based surveillance, which relies on patients to self-report new events. While this type of approach raises concerns about the representativeness of patients agreeing to participate, that may not be a problem since there is a control group of people receiving a different drug, subject to the same selection process. Unless willingness to participate was somehow different among the groups of study subjects being compared, no bias should result. Perhaps more serious, however, is the risk of missing many clinical events by relying on patient initiative to report them, and that the degree of incomplete reporting could easily be related to study group or, alternatively, be sufficiently severe to mask any real result due to nondifferential misclassification.

Finally, pharmacy-based surveillance was used to conduct a massive study of parenteral ketorolac (see Case Example 13.1). A retrospective cohort study between November 18, 1991 through to August 31, 1993 identified subjects from 35 hospitals in the Delaware Valley Case-Control Network. Included were 9907 inpatients given 10 279 courses of parenteral ketorolac and 10 248 inpatients given parenteral narcotics and no parenteral ketorolac, matched on hospital, admission service, and date of initiation of therapy. Patients were enrolled by identifying users of these drugs, from the hospital pharmacies. The source of data on these patients was then chart review, using computer-assisted chart abstracting forms. The study concluded that the adverse event profiles of ketorolac and narcotics appeared different, mostly in the pattern predicted, with ketorolac having an increased risk of gastrointestinal bleeding, especially in the elderly, associated with higher doses and with use greater than 5 days, and with narcotics having a higher risk of respiratory depression, but without a difference in risk in many other outcomes. Overall, the risk/benefit balance of parenteral ketorolac versus parenteral opiates was deemed to be similar, but that improving the use of ketorolac (e.g., duration <5 days) would improve risk/benefit balance further, and that the choice of the optimal drug needs to be made on a patient-specific basis. Extensive changes were made to the drug's labeling in response to these results, and these changes protected the drug's availability in some markets where concerns had been raised.

CASE EXAMPLE 13.1: PHARMACY-BASED SURVEILLANCE OF PARENTERAL KETOROLAC

Background

- Case reports suggested that parenteral use of ketorolac may have been associated with an increased risk of gastrointestinal bleeding. Yet, ketorolac was being used in sick, hospitalized patients.

Question

- Is use of parenteral ketorolac associated with an increased risk of gastrointestinal bleeding?

Approach

- A prospective cohort study was performed, comparing 9907 inpatients given 10 279 courses of parenteral ketorolac to 10 248 inpatients given parenteral narcotics and no parenteral ketorolac, matched on hospital, admission service, and date of initiation of therapy.
- Data were collected by computer-assisted chart review, within 35 hospitals.

Results

- Parenteral ketorolac was associated with an increased risk of gastrointestinal bleeding, especially in the elderly, with a dose–response relationship, but only in those who used the drug for a prolonged time period.
- Parenteral narcotics had a higher risk of respiratory depression.
- Overall, the risk/benefit balance of parenteral ketorolac versus parenteral opiates was deemed to be similar, but limiting the use of ketorolac (e.g., duration <5 days) would improve the risk/benefit balance further.
- The choice of the optimal drug needs to be made on a patient-specific basis.

Strengths

- Concurrent and comparable control group.
- Huge sample size.
- Natural setting.

Limitations

- Risk of selection bias: the different analgesics were used in patients who were somewhat different.
- Expensive, time-consuming study.

Summary Points

- Cohort studies need control groups to place the experience of the study group in perspective.
- In any nonexperimental study, the risk of a selection bias must be considered.
- Outcomes that are relatively uncommon require large sample sizes.

AD HOC COHORT STUDIES

The “traditional” approach to recruiting patients into pharmacoepidemiology cohort studies has been for pharmaceutical manufacturers to use their sales representatives (also known as “detail men”) to solicit physicians to enroll the next few patients for whom they prescribe the drug in question. The physicians then provide follow-up information on the results of this treatment. For example, in the Phase IV postmarketing drug surveillance study conducted of prazosin, the investigators collected a series of over 20 000 newly exposed subjects, recruited through the manufacturer’s sales force. The goal of this study was to quantitate better the incidence of first-dose syncope, which was a well-recognized adverse effect of this drug. As another example, when cimetidine was first marketed there was a concern over whether it could cause agranulocytosis, since it was chemically closely related to metiamide, another H₂ blocker which had been removed from the market in Europe because it caused agranulocytosis. This study also collected 10 000 subjects, using a similar design, and found no cases of agranulocytosis.

Although this is the “standard” approach to this type of study, it suffers from a number of important problems. First, it is extremely expensive. The studies mentioned above cost over \$1 million each, without counting in this cost the considerable time of the pharmaceutical representatives.

Second, these studies did not include any control group. A control group was not necessary for them to provide useful information about the questions they were designed to answer. They were designed to quantitate the frequency of a defined medical event in those who were exposed to the drug, rather than to test hypotheses about whether the drug caused particular outcomes. However, in general, the absence of a control group is a major problem. Without a control group, one cannot determine whether the observed frequency of any medical event is larger or smaller than would have been expected. Thus, one would expect that such studies would provide little new information, despite their cost, and this is what has been observed.

It often would be difficult or impossible to enroll appropriate controls for a new drug in this type of study. For example, no other H₂ blocker was available on the market at the time cimetidine was marketed. When a United Kingdom postmarketing surveillance study was performed that compared users of cimetidine to the next eligible patient in their general practitioners' offices, differences were seen which were likely to be due to the underlying disease for which the cimetidine was being administered, rather than to the cimetidine. Whether or not recruiting a valid control group would be possible, however, it would double the already considerable cost of a study of this type.

Finally, the physicians recruited into a study via a pharmaceutical company's sales representatives are unlikely to be representative of all physicians. In addition, there is no way to monitor whether the physicians select patients who are representative of all their patients, recruit patients sequentially, or even provide complete information on the patients selected. For all of these reasons, there is a considerable potential for biased results.

Thus, although this method continues to be used, this is mainly for its marketing potential rather than for the scientific information it will gather. In fact, some so-called postmarketing surveillance studies that are designed in this way are in fact pure marketing efforts, with no real attempt to gather useful scientific information. These are described to participants as pharmacoepidemiology studies, when in fact they are market-seeding studies. This practice is unfortunate, however, and should be abandoned. In addition to being of questionable integrity, this practice is troublesome, as physicians could become jaded as well as disillusioned with and skeptical about postmarketing surveillance studies in general, jeopardizing future studies which could make important contributions.

OTHER APPROACHES

Other approaches to recruiting patients for cohort postmarketing surveillance studies are more opportunistic. As an example, most of the databases described in earlier chapters can be used to identify individuals exposed to selected drugs (see Chapters 11 and 12).

As another example, in a United Kingdom postmarketing surveillance study, the Prescription Pricing Bureau and local pharmacies in four geographic areas were used to identify individuals prescribed cimetidine by their general practitioners. Controls were selected from the general practitioners' practice file, as the next patient in the file of the same sex, age (by decade), and who had attended the practice within the prior 12 months. Outcome data were collected by

visiting the general practitioner again 15 months later and reviewing his/her records for any intervening care received by the patient.

Finally, other approaches can be considered as well, such as by systematically contacting physicians who are likely to be prescribing the drug. For example, to evaluate cimetidine one could solicit via mail the cooperation of all gastroenterologists. As another example, to evaluate a new vaccine one could contact city health clinics that administer the vaccine, and one could solicit the cooperation of pediatricians.

USING RANDOMIZED CLINICAL TRIALS AS POSTMARKETING SURVEILLANCE STUDIES

For the reasons described in Chapter 2, randomized clinical trials do not have as large a role in postmarketing studies as they do in premarketing studies. They are artificial and raise logistical problems. Perhaps most importantly, however, they are often unnecessary, because of the studies performed premarketing. In fact, however, most studies performed after drug marketing *are* randomized clinical trials. Most of those are intended to address specific questions about drug efficacy, and are conducted as if they were premarketing studies. A few are designed to study drug safety. However, there is a relative lack of postmarketing surveillance randomized clinical trials that take advantage of the fact that they are studying an approved drug.

Specifically, pharmacoepidemiology techniques can be used to conduct postmarketing clinical trials in ways that could be less costly and less artificial. An example is the Group Health Cooperative of Puget Sound's randomized clinical trial comparing the toxicity of microencapsulated versus wax-matrix formulations of oral potassium chloride.

Since these are both FDA-approved products and are theoretically interchangeable formulations of the same active drug, this HMO randomly allocated its pharmacies to dispense either the wax-matrix formulation or the microencapsulated formulation.

As another example, one could conduct a large-scale double-blind randomized clinical trial of two different drugs in the same drug class, using participating prescribers to enlist the patients. The study would be performed by mail. After obtaining patient consent, participating prescribers would use special preprinted prescription pads for the "study drug." These prescriptions would be telephoned, and later mailed, into the coordinating center, which would then express-mail the drug to the patient, who would not have to pay for the drug. Data collection would be performed

by questionnaires mailed to the patients and by obtaining copies of the physicians' medical records. The incentives needed to obtain physician participation should be much less than those used in classical premarketing clinical trials, because of the markedly decreased amount of work being requested. This would be explored within a pilot study to be conducted first.

Thus, postmarketing randomized clinical trials can be conducted in innovative ways that take advantage of the postmarketing setting, rather than simply performing a premarketing randomized trial after marketing. This is uncommonly done, however. Much more is presented in Chapter 20 on the use of randomized clinical trials for pharmacoepidemiology studies.

ADDITIONAL DATABASES USEFUL FOR PHARMACOEPIDEMIOLOGIC RESEARCH

Finally, there are a number of other databases potentially useful for pharmacoepidemiologic research. In addition, new databases are continuously under development. In general these, of course, can be used for either cohort or case-control studies. Most have been used only rarely for pharmacoepidemiology, but could be used more often, especially if expanded.

One of these is the medical record linkage system used in Finland. The Finnish Cancer Registry, Congenital Malformation Register, and Hospital Discharge Register can each be linked to the Register of Persons Entitled to Free Drugs. The particular advantages of this system are that it collects nationwide data and some of the notifications are mandatory. The particular disadvantages are the relatively small population size. Finland's population is only 5.2 million in total. There are now only 1.2 million individuals in the register (406 000 with 100% reimbursement, and 800 000 at 75% reimbursement). Nevertheless, it has been used on occasion for formal analytic research in pharmacoepidemiology. For example, a paper reported a case-control study which did not confirm the initial reports of an association between reserpine and breast cancer. As another example, some epidemiologic studies have been conducted by combining information with data from other registers. For example, the total medication pattern of diabetic patients was studied by identifying diabetic patients through the database maintained by the Social Insurance Institution of Finland—where individuals with specified chronic diseases (about 50 diseases) are entitled to special reimbursement for drug treatment costs—and by linking this information to prescription data in the same database, including all details

and costs of medications prescribed, purchased, and reimbursed. Matched controls were chosen from the population registry, identified by the section of the population *not* entitled to special reimbursement for diabetes and who did not purchase antidiabetic medications.

Another database occasionally being used for pharmacoepidemiologic research is the longstanding Regenstrief Medical Record System. This database contains all laboratory, pharmacy, and appointment information for a network of inner city facilities in Indianapolis, including 5 large hospitals, 44 outpatient clinics, 13 homeless care sites, and the county and state health departments, all of which are located in and around Indianapolis. It has been used for many studies, although only a few pharmacoepidemiology studies.

Although it is a uniquely deep data resource, e.g., in its availability of laboratory data, for the purposes of pharmacoepidemiology studies it suffers from a relatively small population and incomplete ascertainment of outcomes, i.e., patients who go to other facilities in Indianapolis for some of their care will not have the associated care recorded. This means that key exposures or outcomes could be missed, as well as important confounders.

Another database used for pharmacoepidemiologic research, very different from Regenstrief, is IMS America's MediPlus database. IMS HEALTH maintains de-identified, population-based, longitudinal patient databases for the UK, Germany, France, and Austria (IMS Disease Analyzer—Mediplus, formerly MediPlus), the information recorded being dependent on the local health care system. IMS Disease Analyzer—Mediplus UK contains full records from around 125 computerized general practices. There are approximately 560 partners at participating practices, over 2 million patient records (~1 million active) and over 95 million prescriptions. Of active patients (i.e., currently registered with a panel GP), nearly 400 000 have prescription history of more than 10 years. The earliest live data entry was in 1991, histories before that date being summarized. This is a UK medical record database, very similar to the General Practice Research Database described in detail in Chapter 12. One difference is the lack of an attempt to add hospital outcome data to these data, so it is mostly useful for studies of outpatient prescribing patterns and of events that do not result in hospitalization. IMS Disease Analyzer—Mediplus Germany contains patient records from 400 practices (290 GPs and 110 internal specialists) since 1992. There are over 4 million patients and over 66 million issued prescriptions recorded. Approximately 45% of patients in the German database have over 3 years of history although, due to the constraints of the health care system, their records

may be incomplete. The German database also allows access to panels of specialists, including gynecologists, pediatricians, urologists, orthopedic surgeons, ENT doctors, dermatologists, surgeons, and neurologists. For Disease Analyzer—Mediplus France, clinical information on over 1 million patients with approximately 22 million prescriptions is collected from around 450 GPs working in a computerized environment. In-depth information on the management of a patient's diagnosis and its treatment can be obtained. In Austria, the database contains about 100 practices (GPs and internists), almost 500 000 patients, and over 11 million prescriptions in total. Though all four databases are primary care-based, participating doctors themselves record hospital admission, specialist referral, and other information that is necessary for a complete clinical record, such as laboratory tests (differences according to country). The databases are increasingly used in research for publication.

Yet another, very different, type of database that has been available for a number of years is the Health Evaluation through Logical Processing (HELP) System at LDS Hospital, Salt Lake City, Utah (540 beds). This is a computerized hospital information system designed to provide administrative, financial, and clinical hospital services. There is patient-based information on the drugs administered, which can be linked to events, other therapies, and procedures. HELP is an integrated expert system that conducts evaluations and provides recommendations on the care of each patient. HELP collects information from admissions, diagnostic laboratories, the pharmacy, nursing care documentation, and more, by cross-checking and analyzing critical information from various departments of Intermountain Healthcare Hospitals. The epidemiologic applications of the HELP system at first have been primarily in the area of infectious diseases and antibiotic use, and have been expanded to other areas as well. Another application of the HELP system has been for computerized surveillance of adverse drug events. The adverse drug event monitoring system combined "enhanced" voluntary reporting by hospital personnel through entry of potential adverse events at any computer terminal in the hospital, with automated detection of adverse drug events through signal events, e.g., sudden discontinuation of a drug, an order for an antidote, certain laboratory tests, and abnormal results. Using this system, 36 653 patients were monitored over an 18-month period. Of these patients, 648 experienced 731 adverse drug events. Before initiation of this program, only 10–20 adverse drug events were reported on a voluntary basis annually. The investigators were able to determine the patient populations at risk for an adverse drug event and those drug classes most often associated with an adverse event.

A number of databases are emerging from Italy as well. There is one in particular, from the Italian region of Friuli Venezia Giulia (FVG), which has been productive of a number of papers in the international literature. All Italian residents are registered with the national health service, which provides free medical care (including hospitalizations) through all of the public and most of the private providers. FVG is a region in the northeastern part of Italy, at the border with Slovenia and Austria. The region is divided into four provinces, with a total of 1.2 million residents with age and sex distributions comparable to those of the rest of Italy. Since 1976 the Regional Directorate of Health in FVG has developed a large automated database (Sistema Informativo Sanitario Regionale, SISR), in which data on all hospitalizations and prescriptions filled in the region are collected, in addition to demographic information and a variety of specialized medical and administrative files.

Drug data have been collected since 1991 or 1992, depending on the province within FVG. The *hospital record file* has accumulated information about all hospitalizations in public and private hospitals within the region since 1985. Since January 1, 1998 an *outpatient file* has collected information on encounters with specialists (including contacts with in-hospital emergency care), laboratory tests, and diagnostic and treatment procedures reimbursed by the national health service.

CONCLUSIONS

In summary, there are a number of other approaches to pharmacoepidemiology studies, in addition to the ones described in detail earlier. No approach to pharmacoepidemiology studies is ideal. Each available approach has its advantages and its disadvantages. In the next chapter, we will place all of these options in perspective, discussing how one chooses among the available alternatives.

Key Points

- Analyses of secular trends examine trends in an exposure and trends in a disease and explore whether the trends coincide, usually using one of many sources of drug sales data and one of many sources of vital statistics data.
- Disease registries can be useful means of case finding for case-control studies.

(Continued)

- *Ad hoc* case-control studies can be very useful for studying uncommon conditions, but the choice of the control group needs to be made carefully, preferably choosing population-based cases and controls.
- *Ad hoc* cohort studies can be useful to study new drugs, but are expensive and large, and need to have a concurrent control group to avoid attributing all adverse events to the drug.
- There are many other approaches to conducting pharmacoepidemiology studies which are discussed briefly in this chapter.

SUGGESTED FURTHER READINGS

- Aromaa A, Hakama M, Hakulinen T, Saxen E, Teppo L, Idan-Heikkilä J. Breast cancer and use of rauwolfia and other antihypertensive agents in hypertensive patients: a nationwide case-control study in Finland. *Int J Cancer* 1976; **18**: 727–38.
- Borden EK, Lee JG. A methodologic study of postmarketing drug evaluation using a pharmacy-based approach. *J Chron Dis* 1982; **35**: 803–16.
- Classen DC, Burke JP. The computer-based patient record: the role of the hospital epidemiologist. *Infect Control Hosp Epidemiol* 1995; **16**: 729–36.
- Colin-Jones DG, Langman MJS, Lawson DH, Vessey MP. Cimetidine and gastric cancer: preliminary report from postmarketing surveillance study. *BMJ* 1982; **285**: 1311–13.
- Colin-Jones DG, Langman MJS, Lawson DH, Vessey MP. Postmarketing surveillance of the safety of cimetidine: 12 month mortality report. *BMJ* 1983; **286**: 1713–16.
- Dawber TR. *The Framingham study. The Epidemiology of Atherosclerotic Disease*. Cambridge, MA: Harvard University Press, 1980.
- Fisher S, Bryant SG, Kent TA. Postmarketing surveillance by patient self-monitoring: trazodone versus fluoxetine. *J Clin Psychopharmacol* 1993; **13**: 235–42.
- Gifford LM, Aeugle ME, Myerson RM, Tannenbaum PJ. Cimetidine postmarket outpatient surveillance program. *JAMA* 1980; **243**: 1532–5.
- Graham RM, Thornell IR, Gain JM, Bagnoli C, Oates HF, Stokes GS. Prazosin: the first dose phenomenon. *BMJ* 1976; **2**: 1293–4.
- Jick H, Jick SS, Walker AM, Stergachis A. A comparison of wax matrix and microencapsulated potassium chloride in relation to upper gastrointestinal illness requiring hospitalization. *Pharmacotherapy* 1989; **9**: 204–6.
- 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.
- McDonald CJ, Tierney WM, Overhage JM, Martin DK, Wilson GA. The Regenstrief Medical Record System: 20 years of experience in hospitals, clinics, and neighborhood health centers. *MD Comput* 1992; **9**: 206–17.
- Melton LJ. History of the Rochester Epidemiology Project. *Mayo Clin Proc* 1996; **71**: 266–74.
- Schlesselman JJ. *Case-Control Studies—Design, Conduct, Analysis*. New York: Oxford University Press, 1982.
- Strom BL, Schinnar R. Participants and refusers in a telephone interview about hormone replacement therapy were equally likely to be taking it. *J Clin Epi* 2004; **57**: 624–6.
- Strom BL, Hibberd PL, Stolley PD. No evidence of association between methyl dopa and biliary carcinoma. *Int J Epidemiol* 1985; **14**: 86–90.
- Strom BL, West SL, Sim E, Carson JL. Epidemiology of the acute flank pain syndrome from suprofen. *Clin Pharmacol Ther* 1989; **46**: 693–9.
- 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.
- 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.
- Varas-Lorenzo C, Garcia-Rodriguez LA, Cattaruzzi C, Troncon MG, Agostinis L, Perez-Gutthann S. Hormone replacement therapy and the risk of hospitalization for venous thromboembolism: a population-based study in southern Europe. *Am J Epidemiol* 1998; **147**: 387–90.

14

How Should One Perform Pharmacoepidemiology Studies? Choosing Among the Available Alternatives

Edited by:

BRIAN L. STROM

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

INTRODUCTION

As discussed in the previous chapters, pharmacoepidemiology 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 pharmacoepidemiology 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 problems. Chapters 7–13 presented each of the different data collection approaches and data resources that have been developed to perform pharmacoepidemiology studies efficiently, despite 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 PHARMACOEPIDEMIOL- OGY STUDIES

Once one has decided to perform a pharmacoepidemiology 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 important to tailor the choice of pharmacoepidemiology 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 pharmacoepidemiology approaches, examining the comparative characteristics of each (see Table 14.1). One can then examine the characteristics of the research question at hand, in order to choose the pharmacoepidemiology approach best suited to addressing that question (see Table 14.2). The assessment and weights provided in this discussion and in the accompanying tables are arbitrary. They are not being represented as a consensus of the pharmacoepidemiology 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 decisions inherent in it, will make it easier for the reader to synthesize the large amount of information presented in the prior chapters.

COMPARATIVE CHARACTERISTICS OF PHARMACOEPIDEMIOLOGIC DATA RESOURCES

Table 14.1 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, The Netherlands Automated Pharmacy Record Linkage System, and Prescription-Event Monitoring in the UK cover entire countries or large fractions thereof. Medicaid databases are next largest, with UnitedHealth Group approaching that, with over 16 million persons. The UK General Practice Research Database (GPRD) has a population of about 3 million individuals. Then, Kaiser in Northern California currently includes 2.8 million subscribers, Kaiser in Southern California about 2 million, and Kaiser Northwest about 440 000. The Saskatchewan database includes about 1 million currently active individuals. 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 for the study at hand.

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 purposes. 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 purposes. 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. Prescription-Event Monitoring in the UK, the health databases in Saskatchewan, the Automated Pharmacy Record Linkage in the Netherlands, and the Tayside Medicines Monitoring Unit in Scotland 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 skewed by the various selection criteria plus their willingness to volunteer for the study. The GPRD uses a nonrandom large subset of the total UK population, and so may be representative. The Group Health Cooperative, the HMO Research Network, Kaiser Permanente, and UnitedHealth include health maintenance organization (HMO) populations. These 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. Some of the remaining data collection approaches or resources are characterized in Table 14.1 as “variable,” meaning their representativeness depends on which hospitals are recruited into the study. *Ad hoc* studies are listed in Table 14.1 “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, all of whose medical

Table 14.1. Comparative characteristics of pharmacoepidemiologic data resources^a

Pharmacoepidemiology approach	Relative size	Relative cost	Relative speed	Representativeness	Population-based	Cohort studies possible	Case-control studies possible
Spontaneous reporting	+++	+	+++	++	—	—	+
Group Health Cooperative	+	+	+++	+++	+	+	+
Kaiser Permanente Medical Care Programs	++	+	+++	+++	+	+	+
The HMO Research Network	+++	+	+++	+++	+	+	+
UnitedHealth Group	+++	+	+++	+++	—	+	+
Medicaid databases	+++	+	+++	+	+	+	+
Health databases in Saskatchewan	+++	+	+++	+++	+	+	+
Netherlands	+++	+	+++	+++	+	+	+
Tayside	+	+	+++	+++	+	+	+
GPRD	+++	+	+++	+++	—	+	+
<i>Ad hoc</i> analyses of secular trends	As desired	+	+++	As desired	—	—	—
Case-control surveillance	Variable	++	+	Variable	—	—	+
Registry-based case-control studies	Variable	++	++	Variable	Variable	—	+
<i>Ad hoc</i> case-control studies	As desired	++	+	As desired	Variable	—	+
Prescription-Event Monitoring	++++	++	++	++++	—	+	—
Pharmacy-based surveillance: outpatient	+++	++	++	Variable	—	+	—
<i>Ad hoc</i> cohort studies	As desired	+++	—	As desired	—	+	—
Pharmacoepidemiology randomized trials	As desired	++++	—	—	—	+	—

Table 14.1. (Continued)

Pharmacoepidemiology approach	Validity of exposure data	Validity of outcome data	Control of confounding	Inpatient drug exposure data	Outpatient diagnosis data	Dates of available data	Loss to follow-up
Spontaneous reporting Group Health Cooperative	++ ++	++ + to +++	— +++	+ +	++ —	Fall 1969 to date 1972 to date	N/A 15%/y × 2y, then 5%/y 3%/y after 2y
Kaiser Permanente Medical Care Programs	++	+ to +++	++	—	—	— ^b	14%/y
The HMO Research Network	++	+ to +++	++	+/-	++	Varies	Unknown
UnitedHealth Group	+	+ to +++	++	—	++	1990 to date	20–25%/y
Medicaid databases	++	+ to +++	++	—	++	— ^c	Nil
Health databases in Saskatchewan	++	+ to +++	++	—	+	9/75–6/87, 1/89 to date	Nil
Netherlands	++	+ to +++	+	—	—	1987 to date	Nil
Tayside	++	+ to +++	++	—	+	1989 to date	Nil
GPRD	++	++ to +++	++	+/-	++	1990 to date	5%
<i>Ad hoc</i> analyses of secular trends	Variable	Variable	—	Variable	+	As desired	N/A
Case-control surveillance	+ to ++	+++	++	—	+	1975 to date	N/A
Registry-based case-control studies	+ to ++	+++	++	—	Variable	Variable	N/A
<i>Ad hoc</i> case-control studies	+ to ++	+++	++	+	+	As desired	N/A
Prescription-Event Monitoring	++	+++	++	—	++	Variable	25%
Pharmacy-based surveillance: outpatient	++	+++	++	—	+	As desired	Variable
<i>Ad hoc</i> cohort studies	++	+++	++	+	++	As desired	Variable
Pharmacoepidemiology randomized trials	++	+++	+++	+	+++	As desired	Variable

^a See the text of this chapter for descriptions of the column headings, and previous chapters for descriptions of the data resources.^b Varies by Kaiser site.^c Varies by state and data vendor.

Table 14.2. Characteristics of research questions and their impact on the choice of pharmacoepidemiologic data resource^a

Pharmacoepidemiology approach	Hypothesis generating ^b	Hypothesis strengthening ^c	Hypothesis testing ^d	Study of benefits (versus risks)	Incidence rates desired	Low incidence outcome
Spontaneous reporting	++++	+	—	+	—	++++
Group Health Cooperative	+	++++	+++	+	+++	+
Kaiser Permanente Medical Care Programs	++	+++	+++	+	+++	+++
The HMO Research Network	+	++++	+++	+	+++	+++
UnitedHealth Group	+	++++	+++	+	+++	+++
Medicaid databases	++	++++	+	+	+++	++++
Health databases in Saskatchewan	+	++++	+++	+	+++	++
Netherlands	+	++	++	+	+++	++++
Tayside	+	++++	++	+	+++	+
GPRD	+	++++	+++	+	+++	++
<i>Ad hoc</i> analyses of secular trends	+	+++	—	+	—	++++
Case-control surveillance	++	+++	+++	+	—	+++
Registry-based case-control studies	+	+++	++	+	+	+++
<i>Ad hoc</i> case-control studies	+	++	+++	+	+	+++
Prescription-Event Monitoring	+	++	+++	+	++	++++
Pharmacy-based surveillance: outpatient	+	++	+++	+	++	++
<i>Ad hoc</i> cohort studies	+	++	+++	+	++	++
Pharmacoepidemiology randomized trials	+	+	++++	++++	+++	+

Pharmacoepidemiology approach	Low prevalence exposure	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	++++	+	++++	++++	+	+	++++	++++
Group Health Cooperative	+	+++	—	—	—	+++	++	++++
Kaiser Permanente Medical Care Programs	+++	+++	—	+	—	++	++	+++
The HMO Research Network	+++	+++	+ / —	+	—	+	++	++
UnitedHealth Group	+++	++	—	+++	—	—	++	+++
Medicaid databases	++++	+	—	+++	—	+	+++	+++
Health databases in Saskatchewan	++	++	—	++	—	++++	++	+++
Netherlands	++++	+	—	—	—	++	+++	+++
Tayside	+	+	—	—	—	—	+++	+++
GPRD	++	++	+ / —	++++	—	—	++	+++
<i>Ad hoc</i> analyses of secular trends	—	—	—	—	—	+	+	++
Case-control surveillance	+	+++	—	—	—	++	+	+ to ++++
Registry-based case-control studies	+	++	—	+	+	++	+	+
<i>Ad hoc</i> case-control studies	+	+++	++++	++++	—	++	+	+
Prescription-Event Monitoring	++++	++	—	++++	++	+	++++	+ to ++++

Table 14.2. (Continued)

Pharmacoepidemiology approach	Low prevalence exposure	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
Pharmacy-based surveillance: outpatient	+++	+++	—	++++	++	+	++++	+
<i>Ad hoc</i> cohort studies	+++	+++	+	++++	++	+	++++	+
Pharmacoepidemiology randomized trials	+++	++++	+	++++	++++	+	++++	+

^a See the text of this chapter for descriptions of the column headings, and previous chapters for descriptions of the data resources.

^b Hypothesis-generating studies are studies designed to raise new questions about possible unexpected drug effects, whether adverse or beneficial.

^c Hypothesis-strengthening studies are studies designed to provide support for, although not definitive evidence for, existing hypotheses.

^d Hypothesis-testing studies are studies designed to evaluate in detail hypotheses raised elsewhere.

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. 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 which 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, nonparticipating, hospital if the disease recurred. This recurrence would not be detected in a study using such a system. The data resources that are population-based are those that use data from organized medical 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, but this is unusual (see also Chapters 2 and 13).

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.

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 compliance in these data. Exposure data from organized systems of medical care 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 compliance. In addition, in HMOs there are drugs that may fall beneath a patient's deductibles, or not be on formularies. For UnitedHealth Group, since Medicare drug benefits vary depending on the plan, pharmacy files may not capture all prescribed drugs if beneficiaries reach the drug benefit limit. In the GPRD, drugs prescribed by physicians other than the general practitioner could be missed, although continuing prescribing by the general practitioner would be detected. 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. However, this would be expected to vary, depending on the type of drug taken, the questioning technique used, etc. (see Chapter 15).

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 GPRD analyzes the actual medical record, rather than claims, and can access additional questionnaire data from the general practitioner as well.

Control of confounding refers to the ability to control for confounding variables. The most powerful approach to controlling for confounding is the randomized clinical trial. As discussed in Chapter 2, the randomized clinical trial is the only 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 Group Health, GPRD, Kaiser, case-control surveillance, *ad hoc* case-control studies, and *ad hoc* cohort studies. The health databases in Saskatchewan, UnitedHealth Group, Tayside, Medicaid (sometimes), and the HMO Research Network can obtain primary medical records, but not all information necessary is always available in those records. They generally are unable to contact patients to obtain supplementary information that might not be in a medical record. Medicaid databases have considerable additional data available, but are no longer as certain to be able to access medical records. Finally, spontaneous reporting systems and analyses of trends 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, Group Health (since 1989), Harvard Pilgrim Health Care (within the HMO Research Network), *ad hoc* studies, and the rare analyses of trends designed to study inpatient drug use.

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 GPRD, the HMO Research Network, Prescription-Event Monitoring, and some *ad hoc* cohort studies. Other outpatient data come via physician claims for medical care, including Medicaid databases, UnitedHealth Group, and the health databases in Saskatchewan. The latter include outpatient data, but only to three digits of the ICD-9 coding

system. 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 by the patient. These include most case-control studies and outpatient pharmacy-based monitoring.

The *start dates and duration* of the available data differ substantially among the different resources, as does the degree of *loss to follow-up*. They are specified in Table 14.1.

CHARACTERISTICS OF RESEARCH QUESTIONS AND THEIR IMPACT ON THE CHOICE OF PHARMACOEPIDEMIOLOGIC DATA RESOURCES

Once one is familiar with the characteristics of the pharmacoepidemiology resources available, one must then examine more closely the research question, to determine which resources can best be used to answer it (see Table 14.2).

Pharmacoepidemiology 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 approaches that have attempted to do this systematically have been Kaiser Permanente, case-control surveillance, Prescription-Event Monitoring, and Medicaid databases, none of which have resulted in notable new findings. To date, the most productive source of new hypotheses about drug effects has been spontaneous reporting.

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 data set, evaluating a hypothesis that arose elsewhere. Because potentially confounding variables would not 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. 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 exceptions—see Strom *et al.*, 1989). Analyses of trends cannot be used to test hypotheses as they cannot control for confounding. 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. Techniques which allow access to patients and their medical records are the next most powerful, as one can gather information on potential confounders that might only be reliably obtained from one of those sources or the other. Techniques 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 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 21. As described there, it is sometimes possible to address these questions using nonexperimental study designs. Generally,

however, the randomized clinical trial is far preferable, when feasible.

In order to address questions about the *incidence of a disease* in those exposed to a drug, one must be able to quantitate 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 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 3). This can best be done using spontaneous reporting, Prescription-Event Monitoring, the Netherlands system, or *ad hoc* analyses of secular trends, which can or do cover entire countries. Alternatively, one could use UnitedHealth Group, the HMO Research Network, or Medicaid databases, which cover a large proportion of the United States, or the GPRD in the UK. Kaiser in Northern California includes 2.8 million subscribers, Kaiser in Southern California over 2 million, and Kaiser Northwest about 440 000. Saskatchewan contains a population of about one million. Pharmacy-based surveillance methods and *ad hoc* cohort studies could potentially be expanded to cover equivalent populations. The Group Health Cooperative includes fewer individuals and so would be less useful to answer questions about uncommon outcomes. Tayside is also small. 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, but this would be very difficult and costly.

To address a question about a *low prevalence exposure*, one also needs to study a large population (see Chapter 3). Again, this can best be done using spontaneous reporting,

the Netherlands system, or Prescription-Event Monitoring, which cover entire countries. Alternatively, one could use UnitedHealth Group, the HMO Research Network, or Medicaid databases, which cover a large proportion of the United States, or the GPRD in the UK. Pharmacy-based surveillance methods and *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. *Ad hoc* analyses of trends would not be useful, as a change in the prevalence of a rare exposure is unlikely to affect the general burden of disease enough to be detectable.

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 and analyses of trends cannot be used for this purpose. The most powerful approach is a randomized trial, as it is the only way to control for unknown or unmeasurable confounding variables. Techniques which allow access to patients and their medical records are the next most powerful, as one can gather information on potential confounders that might only be reliably obtained from one of those sources or the other. Techniques which allow access to primary records but not the patient are the next most useful.

If the research question involves *inpatient* drug use, then the data resource must obviously be capable of collecting data on inpatient drug exposures. The number of approaches that have this capability are limited, and include spontaneous reporting systems, Harvard Pilgrim Health Care (part of the HMO Research Network), and inpatient pharmacy-based surveillance 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 GPRD, 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 can also be used, as they include outpatient data, although one must be cautious about the validity of the diagnosis information in outpatient claims. Saskatchewan is similar, although outpatient data is more limited. 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. 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.

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 is the health databases in Saskatchewan. Drug data are available 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. Group Health Cooperative, Kaiser Permanente, and parts of the HMO Research Network have even longer follow-up times available. However, as HMOs they suffer from some turnover, albeit more modest after the first few years of enrollment. Analogously, 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 *ad hoc* analyses of secular trends, Prescription-Event Monitoring, outpatient pharmacy-based surveillance, *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 three years after a drug is marketed. The major databases are the next most useful, especially Medicaid databases, the HMO Research Network, and UnitedHealth, as their large population bases 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. *Ad hoc* analyses of secular trends and case-control studies, by whatever approach, must wait until sufficient drug exposure has occurred to 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. Analyses of secular trends can be mounted faster than other *ad hoc* studies, and so these can be useful sometimes when an alternative approach will not work. 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, or Tayside). Others could be used to explore inpatient or outpatient bleeding (e.g., Medicaid, Saskatchewan). 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 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 outpatient pharmacy-based surveillance 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 that has been on the market for, say, five years is a cause of a relatively rare condition, such as allergic hypersensitivity reactions. Because of the infrequency of the disease, one would need to draw on a very large population. The best alternatives would be Medicaid databases, the HMO Research Network, UnitedHealth Group, *ad hoc* analyses of trends, case-control studies, or Prescription-Event Monitoring. To

expedite this hypothesis-testing study and limit costs, it would be desirable if it could be performed using existing data. Prescription-Event Monitoring and case-control surveillance would be excellent ways of addressing this, but only if the drug or disease in question, respectively, had been the subject of a prior study. Other methods of conducting case-control studies require gathering exposure data *de novo*.

As a last example, one might want to follow up on a signal generated by a spontaneous reporting system, designing a study to investigate whether a drug which has been on the market for, say, three years is a cause of an extremely rare but serious illness, such as aplastic anemia. One's considerations would be similar to those above, but even Medicaid databases would not be sufficiently large to include enough cases. 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 pharmacoepidemiology 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 pharmacoepidemiology 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.

Key Points

- There are many different approaches to performing pharmacoepidemiology studies, each of which has its advantages and disadvantages.
- It is important to tailor the choice of pharmacoepidemiology resources to the question to be addressed.
- One may want to use more than one data collection strategy or resource, in parallel or in combination.

- By considering the characteristics of the pharmacoepidemiology 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.

SUGGESTED FURTHER READINGS

- 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.
- Coulter A, Vessey M, McPherson K. The ability of women to recall their oral contraceptive histories. *Contraception* 1986; **33**: 127–39.
- Glass R, Johnson B, Vessey M. Accuracy of recall of histories of oral contraceptive use. *Br J Prev Soc Med* 1974; **28**: 273–5.
- Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformations. *Am J Public Health* 1967; **57**: 2071–5.
- Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.
- Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 1982; **116**: 114–22.
- 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.
- 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.
- Schwarz A, Faber U, Borner K, Keller F, Offermann G, Molzahn M. Reliability of drug history in analgesic users. *Lancet* 1984; **2**: 1163–4.
- 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.
- Strom BL, West SL, Sim E, Carson JL. Epidemiology of the acute flank pain syndrome from suprofen. *Clin Pharmacol Ther* 1989; **46**: 693–9.

Validity of Pharmacoepidemiologic Drug and Diagnosis Data

The following individuals contributed to editing sections of this chapter:

SUZANNE L. WEST,¹ BRIAN L. STROM,² and CHARLES POOLE¹

¹School of Public Health, University of North Carolina at Chapel Hill, North Carolina, USA; ²University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 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. Although this statement referred to questionnaire data, it applies equally to data generated by abstracting medical records or data from automated 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, the methodologic problems involved in validity assessment are presented, with some background on measurement error and the most recent information on the cognitive theories of memory systems, including semantic and episodic memory. The chapter then describes selected studies that have evaluated the validity of drug, diagnosis, and hospitalization data and the factors that influence the accuracy of these data. Case studies are provided as illustrative examples of topics, studies, and

databases covered more comprehensively in Chapter 45 of *Pharmacoepidemiology* (4th edition, Brian L. Strom, editor). The chapter concludes with a summary of our current knowledge in the field as well as directions for future research.

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 (see Case Example 15.1). The patient needs to recall the physician's instructions for most efficacious 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. Patient recall may be even poorer for illnesses and medication use that occurred many years previously.

CASE EXAMPLE 15.1: HOW WELL CAN PATIENTS DESCRIBE CURRENTLY USED MEDICATIONS?

Background

- Patients may be treated concurrently or sequentially by several different physicians, all of whom maintain their own office records.
- Any one physician may not have a complete treatment history and may need to rely on the patient to provide this information, especially for drugs that were not efficacious or that resulted in an adverse drug reaction.

Question

- If patients are asked about their medications during an office visit, how well do they recall the names, doses, and frequency of use for the medications they currently take?

Approach

- A survey was conducted of 774 patients attending a university cardiology outpatient clinic in Glasgow.

Results

- 15% of patients attended with their medications, 19% brought a comprehensive medication list, 9% brought a list of the medication names only, 40% were confident that they knew their medication regimens but did not bring their medications with them, and 17% were unsure of the medications they were on. The 17% who were unsure of what they were taking dropped to 2% once appointment cards indicated that the patients should bring their medications to their clinic visits.

Strengths

- A six-item questionnaire was designed to elicit whether subjects either brought all their medications to the clinic or could recall the names, amounts, and regimens of the medications they were currently taking in a non-judgmental manner.
- Data were collected when the subjects registered for clinic, prior to interaction with clinical staff.

Limitations

- Study carried out at only one specialty clinic.
- For those who did not bring their medications to clinic, analysis did not assess accuracy of recall by comparing the drugs in the medical record with those the subject thought he/she was taking.

Summary Points

- A good number of patients cannot communicate their medication regimens to the clinician, which can impede their care.
- A simple reminder stamped on the appointment card can provide physicians with knowledge that will improve the care of their patients.

Of particular concern to the subject of this textbook is the validity of data on drug exposure and disease occurrence because the typical focus of pharmacoepidemiologic research is often the association between a medication and an adverse drug event. Further, many potential confounders of importance in pharmacoepidemiologic research (although certainly not all) are either drugs or diseases. 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 without 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 pharmacoepidemiology 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 ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

COGNITIVE THEORY OF AUTOBIOGRAPHICAL MEMORY

Epidemiologic research often relies on asking study subjects to recall events or exposures that occurred at some time in the past, with recall intervals spanning from days to several years. To appreciate the accuracy of data derived

by recollection, it is important to 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. Often, too little attention is paid to the first two key tasks when developing survey instruments, the result of which is questions that are too vague or complex for respondents to marshal retrieval processes appropriately. Thus, understanding memory organization and retrieval are critical components for developing questionnaires for collecting accurate health-related data on drug use or events such as doctor visits.

Based on current memory theory, autobiographical memory is used to store events, most of which are catalogued for retrieval in a more general fashion and are sequenced based on important personal milestones. Thus, when respondents are asked to recall a visit to a doctor that may have occurred at a particular point in time, researchers believe that the respondents use scripts (a generic mental representation of the event) to help retrieval. For example, the respondent first contemplates a doctor visit in general and then supplements this script with details relevant to the particular visit, which requires contemplation for specific criteria (e.g., diagnosis) and timing (e.g., a particular year).

Autobiographical memory appears to contain generic events that are somehow linked to details for individual events that occurred during specific points in a person's lifetime. How autobiographical memory is organized is still being debated by memory theorists. Conway suggests that autobiographical memory is based on three levels, where the highest level relates to periods defined by personal life events such as one's first job, first child, first year of college, etc. This period organization has both a temporal and thematic structure so that subsequent events can be catalogued appropriately. The next level stores general knowledge and script-type generic information whereas the third level contains detailed information to distinguish among events. This third level, which is more sensory and perceptual, is thought to be similar to episodic memory, described below.

There are four types of temporal questions often included in questionnaires. They include:

- 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?";
- 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?"

The cognitive processes the respondent uses to answer temporal questions will determine the accuracy of the response, and respondents use many different recall processes in combination to develop a response. Memory researchers believe that, although the recall process is individualistic, the types of information used and its integration typically include four concepts, namely recalling the exact date, temporal sequencing, correlating with other events, and determining adequacy of response.

An example best illustrates the theory of Tourangeau and colleagues 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 1998: The recall process begins with the respondent being uncertain whether the depression was diagnosed in 1997 or 1998. 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-1997, just as their children finished the school year. So, it was 1997 when he lost his job, toward 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 1998? It was January 1998 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 is diagrammed in Figure 15.1.

Landmark events such as marriage and childrearing probably serve as the primary organizational units of autobiographical knowledge and, as such, provide an anchor for information retrieval. In particular, the example shows how the respondent used landmark and other notable events, relationships among datable events, and general knowledge

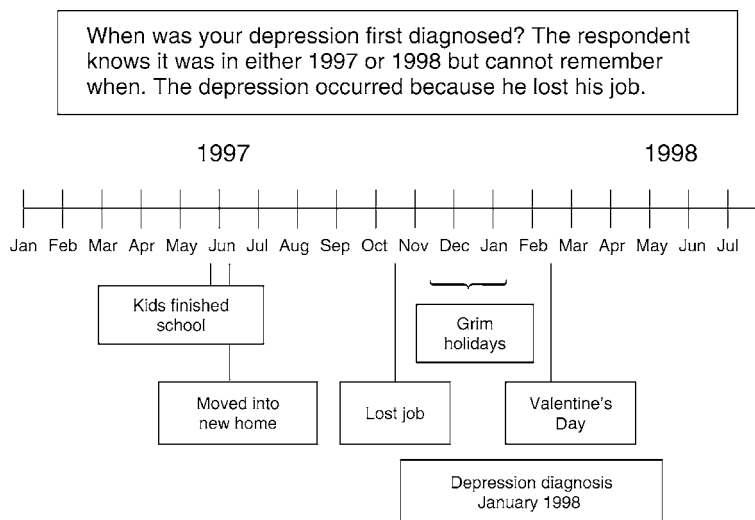


Figure 15.1. Recall process for diagnosis of depression.

(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.

THE INFLUENCE OF MEASUREMENT ERROR ON 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, whether due to inaccurate recall or poorly acquired data, may identify a risk factor in the study which 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. In a study of 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 nonexposed. This misclassification is a measurement error. Although the measurement process usually 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: nondifferential and differential (see also Chapter 16).

The difference between these errors relates to the variables under study. In particular, differential misclassification is said to occur 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 can be toward or away from the null. For example, in a case-control study of oral contraceptives (OC) and breast cancer, there would be concern that those with breast cancer would recall past OC use differently from those without breast cancer. Cases might ponder the origins of their illness and recall and report OC use they otherwise would have forgotten or failed to report. Alternatively, cases might be distracted by their illness during the interview and forget their past OC use, or 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 that disease (e.g., pesticide exposure). The state of mind of the respondent and of the interviewer at the time of the interviews are crucial determinants of the overall accuracy of the interview or questionnaire information and of 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 questioning how these illnesses might have come about. In earlier stages, attention is often directed inward toward self-blame. As the time passes, external explanations are often sought. The time course of the psychological state of seriously ill patients and their close family

members is highly variable, but potentially of great importance to the validity of interview and questionnaire data obtained from them. The traditional assumptions that cases remember true exposures better than non-cases (i.e., that exposure classification has higher sensitivity among cases than among controls) 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 OC exposure and the resulting measure of association between OCs and breast cancer, resulting in recall bias. It is commonly thought that the potential for recall bias can be minimized if the study is designed to obtain complete exposure data, i.e., information on the names and usage dates for every drug used in the time period of interest.

Nondifferential 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 nondifferential misclassification of exposure as well; it is usually biased toward the null. Exceptions can occur when classification errors are not independent of each other, 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 nondifferential misclassification can occur when there are more than two categories of exposure. Rothman and Greenland give a simple hypothetical example to illustrate the potential for bias away from the null from independent, nondifferential misclassification of an exposure with more than two categories of exposure. Consider a case-control study of an exposure with three categories—low, medium, and high—and suppose the correctly classified case and control counts are 100/100, 200/100, and 600/100, respectively. With low exposure as the referent, the odds ratios are 2.0 for medium exposure and 6.0 for high exposure. Now suppose that 40% of the cases and controls in the high exposure group are misclassified into the medium exposure group. The odds ratio for high exposure is unbiased, $(360/60)/(100/100) = 6.0$, and the odds ratio for medium exposure is biased upward to $(440/140)/(100/100) = 3.1$. Finally, there is no bias from independent, nondifferential misclassification of a binary outcome measure under some circumstances. For instance, if there are no false-positive cases, the expected risk ratio will be the risk ratio given

for 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 nondifferential, this ratio equals unity and the risk ratio is unbiased. In addition, it is important to keep in mind that when an expected bias is toward the null, this is the direction of the bias on average. The actual bias in any given study may be away from the null even when the misclassification probabilities are nondifferential.

QUANTITATIVE INDICES OF MEASUREMENT ERROR

Three kinds of comparisons may be drawn between two (or more) methods of data collection or sources of information on exposure or outcome. Many different terms have been used to describe each of them, resulting in a certain amount of confusion.

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 would be a comparison of responses in repeat interviews using the same interview instrument. Reliability is not validity, though the term is sometimes used 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 or reliable.

Only when one of the methods or sources is clearly superior to the other can the comparison be said to measure validity, a synonym for which 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.

For a binary exposure or outcome measure, such as ever vs. never use of a particular drug, there are two measures of validity. 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).

		Gold standard		
		Exposed	Not exposed	
Questionnaire data	Exposed	A true positive	B false positive	m_1
	Not exposed	C false negative	D true negative	m_2
		n_1	n_2	N

Sensitivity = $A/A + C$

Specificity = $D/B + D$

Figure 15.2. Formulas for calculating sensitivity and specificity.

Figure 15.2 illustrates the calculation of sensitivity and specificity. For a drug exposure, a true gold standard would be 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 kept by the study participants or, perhaps more readily available, a computerized database of filled prescriptions, although neither of these data sources might be a genuine gold standard. Prescription diaries cannot be assumed to be kept in perfect accuracy. For instance, there may be a tendency to record that drug use was more regular and complete than it actually was, or that it was used according to the typical prescribed regimen. Similarly, there may be substantial gaps between when a prescription is filled and when it is ingested.

Sensitivity and specificity are the two sides of the validity coin for a dichotomous exposure or outcome variable. In general, sources or methods that have high sensitivity tend to have low specificity, and methods with high specificity tend to have low sensitivity. In these situations, which are very common, neither of the two sources or methods being compared can be said to have superior overall validity to the other. Depending on particulars of the study setting, either sensitivity or specificity may be the more "important" validity measure. Moreover, absolute values of these measures can be deceiving. For instance, if the true prevalence of ever use of a drug is 5%, then an exposure classification method or information source with 95% specificity (and perfect sensitivity) will 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. 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%.

As measures of validity, sensitivity and specificity have "truth" (i.e., the classification according to a gold standard or an alloyed gold standard) in their denominators. Investigators should take care not to confuse these measures with the predictive values of positive and negative classifications, which have the classification according to 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 ($A/(A+B)$ in Figure 15.2). The proportion of persons classified as lacking the exposure or outcome who are correctly classified is the negative predictive value ($D/(C+D)$ in Figure 15.2). 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 on the true prevalence of the exposure or outcome as well. 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 dispensation 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 any way that would not be recorded in the prescription dispensing records).

In general, it is all too common for studies that measure mere agreement to be interpreted as though they measured validity or accuracy. The term “reliability” tends to be used far too broadly, to refer not only to reliability itself, but to agreement or validity as well. A widespread increase in the care with which such terms are used would be very helpful.

There are two methods 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). Realizing that it is only generalizable to populations with similar exposure distributions, the product-moment correlation coefficient may also be used. However, 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 under-estimates values of the other variable. The high correlation means that the over- or under-estimation is systematic and very consistent. When the two measures being compared are plotted against one another 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. However, one is said to have perfect correlation 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.

Quantitative Measurement of Reliability

To evaluate reliability for categorical variables, the percentage agreement between two or more sources and the related (κ) coefficient are used. They are used only when two imperfect classification schemes are being compared, not when there is one classification method that may be considered *a priori* superior to the other. The κ statistic is the percentage agreement corrected for chance. Agreement is conventionally considered poor for a 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. Figure 15.3 illustrates the percentage agreement and κ calculations for a reliability assessment between questionnaire data and medical record information.

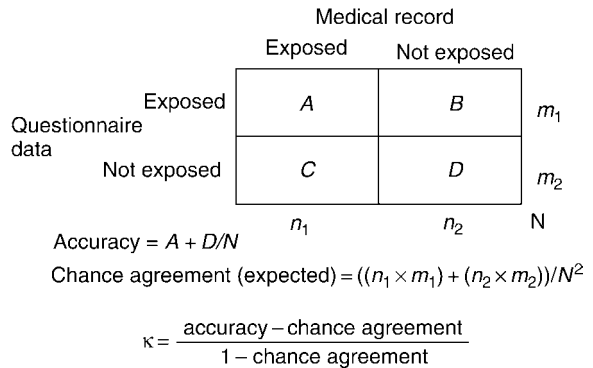


Figure 15.3. Formulas for calculating the percent agreement and κ .

The intraclass correlation coefficient is used to evaluate the reliability of continuous variables. 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 . It is impossible to translate values of measures of agreement, such as κ , into expected degrees of bias in exposure or disease associations.

EFFECTS OF MEASUREMENT ERROR ON THE POINT ESTIMATE OF ASSOCIATION

Copeland *et al.* (1977) evaluated misclassification in epidemiologic studies using a series of computer-generated graphs. They showed that the bias—i.e., 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. 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 . This means that 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 nondifferential misclassification occurred, the point estimate was biased toward the null. Their results

for nondifferential misclassification 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 (1977) were all done on binary disease and exposure variables. Dosemeci *et al.* (1990) presented additional simulations to show that nondifferential misclassification of exposure may bias the point estimate toward or away from the null, or may cause the point estimate to change direction when polychotomous exposure variables (i.e., variables with more than two categories) are considered. A typical example of a polychotomous variable would be never, some, or frequent use of a drug. For a continuous variable, nondifferential misclassification may not produce a bias towards the null if there is perfect correlation between the variable as measured and the true value. For example, if both cases and controls in a case-control study underestimate duration of drug use by an equal percentage, then there would not be a bias towards the null.

CORRECTING MEASURES OF ASSOCIATION FOR MEASUREMENT ERROR

To correct effect estimates for measurement error, estimates of sensitivity and specificity are required. These estimates can be derived from previous research or from a subsample within the study being analyzed. However, estimates of sensitivity and specificity of exposure classification from previous research are rarely available. Should these estimates be available, they may not prove to be useful since the classification methods need to be similar in both the correctly classified and misclassified data. 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.

For differential misclassification of exposure by disease status (recall bias), Raphael (1987) contends that it is the researcher's responsibility to either present a strong case that recall bias did not threaten the study's validity or to control for it statistically. One extremely important way to help make the case for which Raphael has called is to conduct a sensitivity analysis (see also Chapter 16). Sensitivity analysis is the last line of defence 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 analytic methods ranging from the very simple to the highly complex 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. Sometimes it may be found that wildly implausible degrees of inaccuracy would have to have been present to produce observed associations. Other times it may be found that the overall study results would be appreciably biased by values of sensitivity and specificity that, if viewed in isolation and out of the context of the particulars of the study in question, might seem high enough or close enough to being nondifferential to be reassuring.

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 nondifferential misclassification of disease proposed by Horwitz and Feinstein (1978) to explain associations between early exogenous estrogen preparations and endometrial cancer. When proper sensitivity analyses were conducted of this bias, it was found to be capable of explaining only a negligible proportion of those associations.

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. 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 the assessment of selection bias in a study of topical coal tar therapy and skin cancer among severe psoriasis patients; exposure misclassification and selection bias in a study of propanolamine use and stroke; and selection bias, confounder misclassification, and unmeasured confounding in a study of less than definitive therapy and breast cancer mortality, as well as in other clinical and nonclinical applications. Pharmacoepidemiologists would be well advised to continue setting good examples in the application of these methods. 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 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.

CURRENTLY AVAILABLE SOLUTIONS

OVERVIEW OF APPROACHES USED TO EVALUATE THE ACCURACY OF PHARMACOEPIDEMOLOGIC DATA SOURCES

The accuracy of drug exposure and diagnostic data has been measured in pharmacoepidemiology studies, sometimes as a validation effort in etiologic studies and elsewhere as separate methodologic evaluations.

Medical Record Validation for Etiologic Studies

In etiologic studies, where drug exposure and disease occurrence are typically derived from questionnaires, "validation" is often done by comparison with medical records. 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 questionnaire data and medical records, because the medical record itself is not a true "gold standard" for several reasons.

First, retrieval of medical records depends not only on a person's ability to remember and report who prescribed

the drug or diagnosed the condition in question, but on 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.

In addition, even if the medical record is available, it may not list all diagnoses and medications prescribed. The therapeutic class may also affect medical record completeness, with psychotropic medications being poorly documented. Record completeness is likely to vary by type of drug, type of chart (outpatient vs. inpatient), and the number of drugs prescribed in a given period. In summary, the medical record does not document all medications prescribed for individuals. This diminishes the usefulness of medical records for verifying self-reported drug exposure. Of particular concern is that most studies that use medical or pharmacy records to verify exposure are only unidirectional. They confirm drug exposure if reported but typically do not evaluate whether a respondent omits reporting an exposure that actually occurred, i.e., the validation efforts typically assess sensitivity but not specificity.

Methodologic Studies

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, where the rate calculated is a measure of agreement, not a measure of validity. 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 the 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 underreported on questionnaires or are missing due to incompleteness of claims processing in a record-linked database, i.e., data sources with low sensitivity, cannot be 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. In situations where the factor has low prevalence and low sensitivity, a small degree of misclassification can have a dramatic effect on measures of association.

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).

Besides the need for completeness on the individual level, it is also critical that information from all persons who are covered by the health plan from which the database is generated appear in the database. Systematic omissions of specific population groups, such as certain ethnic or racial groups, diminish the quality of the database.

Influences on Accuracy

The accuracy of medication exposure reported via questionnaire is affected by several factors. Research indicates that the type of question influences how well respondents answer medication questions. Recall period, the time between when the exposure occurred and when it is reported, influences accuracy of recall as well (see Case Example 15.2). The extent of use as measured by number of dispensings or duration of use appears to enhance recall.

CASE EXAMPLE 15.2: FACTORS THAT INFLUENCE RECALL ACCURACY

Background

- Much of the pharmacoepidemiologic research in the past 10–15 years has used automated pharmacy claims databases to assess medication exposures.
- However, there are times when the exposure of interest may not be available in these databases because it is an over-the-counter medication or patients pay out of pocket because the copayment is higher than the price for filling the prescription without the copayment.
- In these and other instances, the researcher may have to query the subject to assess medication exposure.

Question

- Is information provided by the subject sufficiently accurate to answer the study question?

Approach

- West and colleagues (1995) conducted a methodologic study to compare subject recall for non-steroidal anti-inflammatory drugs (NSAIDs) and estrogens to

dispensing information from the Group Health Cooperative of Puget Sound (GHC) pharmacy database.

- At the time this study was done, more than 95% of enrollees obtained all of their prescriptions from GHC pharmacies.
- The study was designed to assess whether recall accuracy varied by subject gender, age (50–65 vs. 66–80 years), or timing of use (drugs stopped 2–3 years vs. 7–11 years previously).
- Subjects were selected from the pharmacy database according to these strata and then queried via telephone using a structured questionnaire and memory prompts that consisted of two color documents: one with pictures of brand and generic NSAIDs and the other with brand and generic estrogens.
- Subjects were sent a packet of information that contained an explanatory study letter and a sealed envelope containing the memory prompts. The subjects were told to keep the envelope sealed until the time of the interview. Information on past NSAID and estrogen use was elicited by having the subjects refer to the memory prompts and indicate which medications they used and the timing of use. To further stimulate recall, the subjects were also asked about conditions for which NSAIDs or estrogens were often used, such as arthritis or menopause, respectively.

Results

- The sensitivity for recalling ever use of an NSAID was 57% (95% CI: 50%–64%) and for recalling the estrogen name was 78% (95% CI: 70%–86%).
- Many subjects were unable to recall the names of the NSAIDs they had used in the past and had even more difficulty with the dose, duration, and dates of use.
- Younger subjects had better recall of NSAID use but gender did not influence recall accuracy; age did not influence recall of estrogen use.
- For timing of drug use, the odds of recall were three times higher for NSAIDs and 2.4 times higher for estrogens that were stopped 2–3 years previously compared to those stopped 7–11 years previously.
- The specificity for NSAID or estrogen exposure was 95%.

Strengths

- This study was designed and powered to evaluate differences in recall accuracy by type of medication

(NSAID vs. estrogen), age, gender, and timing of last use.

- The study was able to determine sensitivity and specificity because the pharmacy database was relatively complete and NSAIDs were not approved for over-the-counter use at that time.

Limitation

- Pharmacy dispensings are not indicative of subject adherence; subjects may have filled their prescriptions but never took the medication.

Summary Points

- Subjects have difficulty accurately remembering past medication use and that accuracy differs by age, the type of medication used, and the timing of use.
- Subjects have great difficulty in accurately recalling specific information about past use of NSAIDs but less of a problem recalling past use of estrogens.
- Given that sensitivity and specificity will depend on the medication studied, it is important to realize that the relative importance of differences in these two measures depends on the true prevalence of exposure in the population.

The possibility of recall bias has been the motivation for many validation studies of past medication use. Whereas many of the studies that addressed this issue evaluated recall accuracy for mothers of normal and abnormal infants, others have assessed differences between cases and controls in case-control studies of non-pregnancy related conditions. Overall, the literature does not strongly support a general or uniform indication of recall bias in either circumstance, although there are some exceptions.

To date, few studies have evaluated whether demographic and behavioral characteristics influence the recall of past medication use. Recall accuracy was affected by age in three of five studies evaluated. Few other demographic factors were evaluated consistently across studies. No differences in recall accuracy were noted by gender. Behavioral characteristics such as smoking and alcohol use have been investigated as predictors of recall accuracy. 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. Due to the paucity of information on predictors of recall, further research in this area is warranted.

Conclusions

Most of the work on recall accuracy for past medication use has focused on OCs and replacement estrogens. The results of the OC and replacement estrogen studies indicate that these medications are recalled accurately, especially if researchers allow a range of one year for agreement on age and duration of use. However, women do have difficulty recalling the brands of OC and replacement estrogens used, even if provided with photos or lists of brand names. Recall of nonhormonal drugs appears potentially more problematic. Although we are beginning to see more evaluations of nonhormonal medications, there are still few studies and there are substantial differences among them. For example, recall periods ranged from one month to several years, or the exact number of years was not specified. The drugs studied were as diverse as the recall periods. As more researchers combine verification with their data collection efforts, more information will hopefully become available on the recall accuracy of other types of medications.

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. A cursory review of the recent literature shows that the more recent *ad hoc* case-control studies have focused on medication use just prior to disease onset, a design feature that will improve the validity of medication exposure data. There is also a trend toward using medication-specific and indication-specific questions, along with recall enhancements, which has 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 by no means always, influenced by the type of medication, drug use patterns, the design of the data collection materials, and respondent characteristics. Differential misclassification of exposure by disease status or recall bias, as a major concern in case-control studies of medication use, appears

to be a misapprehension. 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. Many researchers are turning to administrative databases to decrease potential exposure misclassification due to inaccurate self-report.

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. It is difficult to assess the reporting accuracy 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 told so by a clinician. In studies comparing self-reports to clinical evaluation, depending upon the type of condition, there is both under- and over-reporting.

The reporting of a medical condition during an interview is influenced by several factors, including the type of condition as well as the subject's understanding of the problem. Reporting is also dependent upon the respondent's willingness to divulge the information. Conditions such as venereal disease 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. As a result, conditions considered sensitive are likely to be underreported when ascertained by self-report.

Conditions with substantial impact on a person's life are better reported than those with little or no impact on lifestyle. Other factors that influence reporting accuracy of past diagnoses and hospitalizations include the number of physician services for that condition and the recency of services. There is also consensus that the type of surgery is remembered accurately. There has been a thorough evaluation of the influence of demographic characteristics on reporting of chronic illnesses, although the results are conflicting. The most consistent finding is that recall accuracy decreases with age, although this may be confounded by recall interval, or cohort (generational) effects. Whether gender or educational level influences recall accuracy is uncertain. There was a consistent finding that reporting of both illnesses and hospitalizations was better for whites than for non-whites.

As with the validity of medication data, the validity of disease and hospitalization data obtained by self-report is

also influenced by questionnaire design. Providing respondents with a checklist of reasons for visiting the doctor improves recall of all medical visits. This research has also indicated that 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.

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 embarrassing and that do not substantially alter the person's lifestyle are not reported completely. Likewise, reporting accuracy is dependent upon the consistency of the terminology—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.

VALIDITY OF PHARMACOEPIDEMIOLOGIC DRUG AND DIAGNOSIS DATA FROM COMPUTERIZED DATABASES CONTAINING ADMINISTRATIVE OR 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 were reviewed in Chapters 11 and 12. One major advantage of using such databases for pharmacoepidemiologic research is the comparative validity of the drug data in lieu of questionnaire data, where recall bias is always a concern, as previously described.

In general, the 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) (see also Chapter 14). The databases also differ on the number of demographic variables that are available, with all having age and sex, but few having race, occupation, or a measure of health status. Because most plans were developed primarily for reimbursement, they all have relatively complete data on health service use and charges.

The drawbacks and limitations of the data systems discussed below are important to keep in mind. Their most critical limitation for pharmacoepidemiologic research is the manner in which health insurance is covered in the United States, typically through the place of employment. If the employer changes plans, which is often done on an annual basis, or the employee changes jobs, the plan no longer covers that employee or his or her family. Thus, the opportunity for longitudinal analyses is hindered by the continual enrollment and disenrollment of plan members.

Along these lines, the most critical elements in the selection of a database for research are the completeness and validity of the data. Completeness is defined as the proportion of all exposures and/or events 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. 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.

For the data in an administrative database to be considered valid, those 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 would be determined by comparing the database information with other data sources, such as medical records, administrative or billing records, pharmacy dispensings, procedure logs, etc. In a recent review, Rawson *et al.* (1998) described the different validation analyses that could be conducted to evaluate the usefulness of administrative databases for conducting observational studies, using the Saskatchewan Health databases for illustration. These analyses include reviewing the sources noted above, i.e., medical records, billing records, etc., along with the recommendation that pharmacoepidemiologists assess three other factors as well: the consistency between data files within the same system, surrogate markers of disease such as insulin for diabetes, and time-sequenced relationships, i.e., a diagnostic procedure preceding a surgery.

The availability of validation analyses and consistency checks will be discussed briefly for each database, separately for drug and diagnosis data. For all of the pharmacy dispensation databases that will be described, it is important

to realize that none of them can address adherence and drug ingestion, and that over-the-counter medications typically are not included.

An adherence issue that was first raised in the late 1980s but has received more attention recently is that of unclaimed prescriptions, estimated to occur for approximately 2% of all prescriptions. While two-thirds of the unclaimed prescriptions are for new prescriptions, the proportion of unclaimed prescriptions depends on whether or not the medications are essential. One might ask how unclaimed prescriptions might affect the validity and completeness of the 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 this insurance adjustment has been made so there may be a substantial number of prescriptions that we, as researchers, believe were dispensed (and used) but had not been used at all. If there are dispensings in the database that were not picked up, then there is no chance that the individual had the drug exposure and our study would suffer from exposure misclassification. This is an active area of research for reasons of patient adherence and loss of revenue for pharmacies.

Drug Data in Administrative or Medical Record Databases

Group Health Cooperative of Puget Sound

The Group Health Cooperative of Puget Sound (GHC) record-linked database was developed as a medical and administrative information system. Hence its drug data have been considered to be of very high quality for pharmacoepidemiologic research (see Chapter 12).

Kaiser Permanente

As discussed in Chapter 12, 90% of members have prescription drug coverage and members with chronic diseases such as diabetes obtain most of their prescriptions (96.7%) from Kaiser pharmacies. Whether members who do not have chronic diseases fill all of their prescriptions at Kaiser pharmacies is unclear, but for Kaiser Permanente Northern California as many as 15–20% of adult members filled at least some of their prescriptions at non-Kaiser pharmacies.

Harvard Pilgrim Health Care

Approximately 90% of Harvard Pilgrim Health Care members have prescription drug benefits that provide a month's supply of drug for a nominal copayment. Drug data may be missing for the 10% of members without drug benefits, for drugs which cost less than the copayment, or for those who do not submit their drug claim for reimbursement. However, drug exposure can be defined on the basis of either dispensing from affiliated pharmacies or prescribing, as indicated in the encounter records. This is a major advantage, and may permit the identification of drug exposures that otherwise would be missing. No formal evaluations of the completeness of these data have been performed.

UnitedHealth Group

UnitedHealth members are derived from commercial, Medicaid, and Medicare populations (see Chapter 12). The percentage of commercial and Medicaid members who have drug benefits has dropped from 93% in 2000 to 90% currently. Because Medicare pharmacy benefits vary by plan, the completeness of drug exposure data on the elderly is somewhat compromised. Like other health plans with pharmacy copayments, medications that are less expensive than the copayment are likely to be missing from the computerized claims database.

Medicaid

As discussed in Chapter 12, Medicaid databases have been used extensively for pharmacoepidemiologic research, mainly due to the validity and completeness of the drug data. Given that Medicaid covers an indigent population, it is less likely that drugs will be purchased outside of the insurance plan when the copayment may range from \$0.50 to \$5.00. Of course, as with the other databases, there is no way to evaluate adherence with drugs dispensed, other than examining patterns of refills for chronically used medications, nor do Medicaid pharmacy claims data capture over-the-counter medications.

Saskatchewan Health Plan

The Saskatchewan Health Plan has been used extensively for pharmacoepidemiologic research, as was discussed in Chapter 12. There are separate plans within the system, i.e., the Prescription Drug Plan, the Saskatchewan Hospital Services Data, etc., and each plan is responsible for verifying and validating its data. There are a series of checks on each information field on the claim submitted to the drug plan before the claim is approved for payment. These checks include verification that the person was eligible for benefits under the program and that the drug dispensed

was eligible for coverage. The Prescription Drug Plan is remarkably complete: all residents except the 9% who have their prescription drug costs paid by another agency are covered by the plan and individuals without coverage can be excluded from studies.

Dutch System

In the Netherlands, there is almost universal computerization of pharmacy records, enabling the compilation of drug histories as almost all patients are reimbursed for prescription medications (see Chapter 12). Despite policies that encourage competition between pharmacies for reducing drug costs, patients typically remain with one pharmacy. This enhances the longitudinal nature of the medication data. The data are believed to be of high quality for three reasons. First, the computerized dispensing records are subject to financial audit, as they are the basis of reimbursement. Second, individuals have traditionally used only one pharmacy. Third, even though there are economic incentives for identifying cost-effective pharmacy care, these incentives are not so great as to promote pharmacy switching.

Tayside Medicines Monitoring Unit (MEMO)

As explained in Chapter 12, all community prescribing is done by the general practitioners in Scotland. By devising a system for capturing and computerizing general practitioner prescriptions dispensed through community pharmacies, MEMO has developed a prescription drug database. The system is not automated, i.e., the dispensing claims for this pharmacy database are entered manually. There are several checks on the accuracy of the data entry, at the time of assigning the Community Health Index number and after entry of system drug codes, i.e., a proportion of prescriptions are dually entered for quality control.

UK General Practice Research Database (GPRD)

The information for this database is amassed from general practitioners who have agreed to provide data for research (Chapter 12). The computerized drug file for this data source is based on physician prescribing, not pharmacy dispensings. Thus, a person may receive a prescription for a medication but choose not to have it filled—the database would have this person as exposed unless algorithms to deal with adherence are developed. Because this system relies on the prescribing done by general practitioners, specialist-prescribed medications would not be available in the database until the person needed to have their prescription refilled, a responsibility of the person's general practitioner. There are two potential drawbacks for using this

pharmacy database for pharmacoepidemiology: (i) adherence, as drug prescribing does not equate with drug use, and (ii) specialist-prescribed medications are not available in the database unless the specialist provides a consultant letter to the person's general practitioner and the general practitioner enters the information from the letter into the database. However, as patients need to obtain a new prescription each time they need a refill, continuity of medication use can be ascertained.

Diagnoses and Hospitalizations in Administrative Databases

Unlike the drug data in administrative databases, where most researchers are comfortable with data accuracy and completeness, there is considerable concern regarding the inpatient and outpatient diagnoses in these databases. The accuracy of the outpatient diagnoses is more uncertain than the 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, inpatient diagnoses are scrutinized for errors by hospital personnel, monitoring that would not occur typically 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 (see also Chapter 12). 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. It is not clear how health care plans deal with "rule-out" diagnoses, i.e., are they included or excluded from the diagnoses in the physician claims files? In addition, the selection of ICD codes for billing purposes may be influenced by reimbursement standards and patient insurance coverage limitations. 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 Health Maintenance Organizations (HMOs) such as Group Health Cooperative or Kaiser Permanente. Lastly, ICD version changes may produce systematic errors. Case Example 15.3 describes the case procedure used by Chan and colleagues (2003) for validating a diagnosis of acute liver failure or injury potentially attributable to hypoglycaemic agents in a study using the HMO Research Network, which includes data from several of the plans described below.

CASE EXAMPLE 15.3: VALIDATION OF A DISEASE ALGORITHM FOR A CLAIMS DATABASE ANALYSIS

Background

- Based on a methodologic study of the validity of Medicaid claims completed by the Research Triangle Institute in 1984, pharmacy data reflect actual medication dispensing but information on diagnosis, particularly from outpatient visits, may not be accurate. The authors concluded that case definitions for administrative claims database studies should be validated using medical record information.

Question

- How valid are codes for acute liver injury in a claims database?

Approach

- Chan and colleagues (2003) conducted a cohort study using data from the HMO Research Network, to determine the incidence of serious acute liver injury in diabetics treated with hypoglycemic medications.
- Trained medical abstractors reviewed potential cases to eliminate those without some type of liver disease or whose liver disease was diagnosed prior to the time period under study.

Results

- Of the 1287 hospitalization episodes reviewed for subjects hospitalized with a liver diagnosis, the first level review indicated that 106 did not have liver disease and 245 had disease diagnosed prior to April 1, 1997.
- The remaining 936 records were reviewed by an internist with the following results: 172 episodes were for liver conditions not relevant to the study, 526 were for chronic liver disease during the period of interest, and 126 were acute events that could be attributed to causes unrelated to drug exposure.
- The third level review of the remaining 112 records for 91 subjects by three hepatologists who were masked to exposure found that 20 patients did not have liver failure, 12 had chronic liver failure, and 24 had an acute onset that was not attributable to

(Continued)

hypoglycemic agents, leaving 35 patients who had acute liver failure or injury that could not be attributed to causes other than hypoglycemic agents.

Strength

- Very careful, staged review.

Limitation

- Of the 171,264 adults from five participating HMOs that received at least one oral hypoglycemic medication between April 1, 1997 and June 30, 1999, there were only 5.6% ($n = 9604$) who received troglitazone. Use of this drug was linked to severe idiosyncratic hepatocellular injury and was removed from the market in March 2000.

Summary Points

- This study documented how an initial cohort of 1287 hospitalizations due to possible liver disease was reduced to 35 patients with the disorder of interest.
- Had the sequential review not been undertaken, it is very possible that the incidence of acute liver injury or failure due to hypoglycemic agents would have been overestimated by the inclusion of cases that were likely to be attributable to etiologies other than drug exposure.
- Such well-defined case definition and validation is needed for most, if not all, pharmacoepidemiologic studies using administrative claims data.

Group Health Cooperative of Puget Sound

The Group Health Cooperative (GHC) outpatient diagnosis information was discussed in detail in Chapter 12. The inpatient diagnostic database has records for all discharges from GHC-owned hospitals (Chapter 12).

Another file contains outside billing information for all admissions to hospitals not affiliated with GHC, especially emergency admissions. The data from this outside claims file are not incorporated into the inpatient database but are available as an annual utilization data set since June 1989.

Kaiser Permanente

As discussed in Chapter 12, the Kaiser Permanente Medical Care Program (KP) is divided into eight administrative regions, with Northern California KP and the Center for Health Research (KP Northwest) having the oldest

research programs. The plan's clinical and administrative databases are typically used to identify health outcomes for the inpatient and outpatient setting. Researchers using these databases confirm the diagnoses using medical record validation, although the results of the validation may not be published.

Harvard Pilgrim Health Care

Harvard Pilgrim Health Care is nearly unique in that epidemiologic analyses use the same automated records that are used by health care providers to deliver care. Therefore, these records are likely to be more complete than information from databases derived from billing diagnoses only. However, they will also suffer from the problems described above regarding the potential incompleteness of medical records. Researchers at Harvard Pilgrim do compare the information available from the automated medical record with that available as the full-text medical record as part of their quality control during project initiation.

UnitedHealth Group

Combining a diagnosis with a marker drug has become a very common technique in pharmacoepidemiologic research to assess the correspondence between a database diagnosis as indicated by an ICD code and actual occurrence of disease. UnitedHealth Group-affiliated health plans are typically independent practice associations but have also offered gatekeeper or capitated models in addition to their open access or discounted fee-for-service models. The different financial incentives for the varying model structures may affect the completeness of the diagnosis data available in the databases. For example, when billing for reimbursement in capitated plans, the individual's diagnosis may not be provided and, as a result, is not available in the research databases. Alternatively, in a discounted fee-for-service plan, there may be a financial incentive to code diagnoses according to the most profitable reimbursement schedules. In using this data source for research, it would be optimal to restrict the study design to members of one model so that differential incentives and policies do not provide an additional source of potential error when using these databases for conducting observational studies.

Medicaid

As a way to manage the escalating costs of the Medicaid program, capitated programs have been put in place to pay for services to Medicaid beneficiaries. Because providers are paid per person treated rather than paying for each encounter with the provider, some of the encounters, e.g., outpatient visits and hospitalizations, may be missing and which are missing may vary by the plan under study. This

occurs despite the requirement from the Centers for Medicare and Medicaid Services that all encounters be recorded even for those in capitated plans.

Considerable effort has been expended on exploring the accuracy of the diagnoses in these Medicaid claims files. The validity of the discharge diagnoses recorded on outpatient medical charts is much more uncertain. The validity of laboratory-driven diagnoses (e.g., neutropenia) is high. However, for diagnoses which are difficult to make correctly or are defined poorly in the ICD system, the validity is much poorer. For this as well as other reasons, obtaining medical records for Medicaid studies is felt to be mandatory (see Chapter 12).

However, obtaining medical or hospital records to validate diagnoses has become much more difficult as a result of the Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA) of 1996, which went into effect from April 2003. Although researchers can obtain these records legally, they must have the appropriate documentation, which includes a data use agreement from the organization supplying the claims data, a waiver of informed consent from the researcher's institutional review board, and, for health information acquired since April 2003, a waiver of HIPAA authorization from the researcher's institutional review boards. This will be discussed in more detail in the final section, "The future."

Saskatchewan Health Plan

Validation substudies were built into the design of several analyses using this database (see Chapter 12). Overall, there appears to be a very high correlation between the information on the charts and that coded in the hospital services system. However, this is not true for all conditions, and validity should be ascertained with each new condition evaluated.

Dutch System

Through collaboration between the Department of Medical Informatics and the Pharmacoepidemiology Unit of the Erasmus University Medical School, the Integrated Primary Care Information (IPCI) system was established for research. This system consists of computerized patient records from approximately 150 GPs covering 500 000 patients. There are two validation studies that compared inpatient diagnoses from the electronic file to diagnoses in the hospital charts (van der Klauw *et al.*, 1993; Heerdink *et al.*, 1998).

Tayside Medicines Monitoring Unit (MEMO)

The Scottish Morbidity Record (SMR) contains information on all acute inpatient admissions based on the discharge

diagnoses, which are abstracted by trained coding clerks. To maintain high quality and accuracy, the SMR databases are audited frequently by the Information and Statistics Division of the NHS National Services Scotland. Researchers have conducted validation studies of these data, comparing the coded diagnoses with the actual medical chart data. Depending on the diagnosis under study, the validation studies have shown the computerized data to be fairly accurate.

UK General Practice Research Database

As a database derived from the general practitioner's primary medical record, one expects very complete documentation of diagnostic information based on visits to these clinicians. However, before these data could be used for research, there needed to be an assessment of its quality. The first determination of quality was done when the database was initiated, at which time the "up to standard" qualification was designated. Currently, for a practice to be considered up to standard it must record a minimum of 95% of the prescribing and patient encounters that occur in the practice. Although quality evaluations began in 1987, most practices did not receive the "up to standard designation" until 1990 to 1991.

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 for enhancing the quality and credibility of the research. Many studies have included the review of original documents to validate the diagnoses under study, which is especially true for the GPRD.

Evaluating the completeness of the databases is much more difficult as it requires an external data source that is known to be complete. When completeness is assessed, it is typically done for a particular component of a study, such as the effect of drug copayments on pharmacy claims or the availability of discharge letters in the GPRD. There has been little research on the completeness of administrative databases in the past 10 years, with the exception of periodic evaluations of the completeness of the GHC pharmacy database. A study published 20 years ago indicated that pharmacy data from administrative databases were of high quality, and, because claims are used for reimbursing pharmacy dispensings, this should continue to be so today. We recognize that adherence is an issue (see Chapter 25) 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, the databases, are complete and of the highest quality.

THE FUTURE

The methodologies for conducting pharmacoepidemiology studies have shifted over the past 25 to 30 years, from total reliance on studies requiring data collection from individuals, to the extensive use of electronic data from either administrative health claims or electronic medical records. Yet, there is still a need to collect data from individuals, especially for herbals and over-the-counter medications and for inexpensive generic medications that may be purchased outside drug plans. This chapter describes the methodologic work that continues to be published on the most valid approaches for collecting medication data via questionnaires, focusing on question design and recall period. Current evidence suggests that recall accuracy for medications diminishes with time so that researchers should focus on recent medication use to reduce the potential for measurement error. We have also discussed techniques such as sensitivity analysis that can inform the potential magnitude of measurement error when data are collected *de novo*. The availability of these analytic tools will allow investigators to provide quantitative estimates of measurement error when describing the limitations of their research rather than a qualitative comment reflecting the potential for measurement bias.

In contrast with *de novo* data collection for pharmacoepidemiologic research, the availability, use, and richness of electronic data sources have increased exponentially. It was about 20 years ago that we realized how useful administrative claims could be for conducting pharmacoepidemiology studies. With time, we found that careful and systematic evaluation of data accruing for administrative purposes could be used to study drug–disease relationships without having to take the time or spend the 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, notably acute myocardial infarction. Just the occurrence of the procedure code, without the results of the procedure, provided useful knowledge on whether or not the individual had the disease under study.

The improved computer technology that resulted in faster processor speeds and increased storage capacity not only led to the efficient merging of data from multiple health

plans but also to the storage of health care data in an electronic format, i.e., electronic medical records. 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 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 US is slowly moving toward electronic medical records, the paper version of the medical record will continue to be used for some time to come.

With the increasing use of individual health data from databases or medical records for pharmacoepidemiologic and health services research came the concern over data privacy issues. Prior to the passing of the Health Insurance Portability and Accountability Act (HIPAA) of 1996, confidentiality was an issue primarily left within the discretion and control of the researcher's organization. The public relied on the Common Rule as overseen by institutional review boards (IRBs) to provide oversight of the confidential nature of the data and to protect the privacy of research subjects. However, confidentiality and privacy protection were only a small part of the IRB review process.

The HIPAA Privacy Rule protects the use and disclosure of information about the health of individuals that is derived by health care providers during the course of treatment, i.e., individualized medical information from which the identity of persons can be determined (see Chapter 19). The Association of American Medical Colleges has been documenting the effects of HIPAA on biomedical and scientific research and has noted that research activities involving data access and patient recruitment have been adversely influenced by HIPAA implementation (Washington Fax, May 6, 2004). In a statement prepared by the American College of Epidemiology for a November 20, 2003 public hearing of the Privacy and Confidentiality Subcommittee of the National Committee on Vital Statistics on the impact of HIPAA on research, Dr Martha Linet noted problems with access to electronic databases such as Medicare and Medicaid, difficulties in obtaining access to medical records by health providers, and increased difficulties with obtaining patient informed consent.

We are seeing the ramifications of HIPAA implementation in all of our research activities. Chapter 19 provides a detailed description of the Privacy Rule and describes the processes that researchers need to follow to be able to access data useful for pharmacoepidemiologic research. It is clear that the process for acquiring research data has become much

more onerous and time-consuming for researchers and IRBs as well. Currently, we are in a transitional phase where many covered entities, i.e. health care providers and those who house the medical records that researchers need to review or those who maintain large data sets for research, are uncertain about HIPAA requirements. In early 2004, denying researchers access to clinical data was the safest option, given the high financial penalty for breach of confidentiality that is levied solely on the covered entity. More than 2 years after HIPAA implementation, we can look back and see that research is getting done, albeit at a slower pace, as covered entities and investigators undergo a steep learning curve to identify the most efficient ways to meet the HIPAA privacy standards with the least impact on research process and progress.

For the future, the major hurdle we face with regard to data validity for pharmacoepidemiologic research is our ability to obtain medical records to validate health outcomes that are derived from administrative claims. Until experience accrues with how to conduct HIPAA-compliant research, and health care consumers and providers become better informed about the Privacy Rule through education and dissemination programs, the validation of health outcomes may be extremely difficult to accomplish or may become so expensive and time consuming that research costs will escalate tremendously.

Key Points

- The validity of diagnosis and drug use data is a function of two properties: how accurately persons who have medical conditions or use drugs of interest are ascertained (sensitivity); and the accuracy with which those who do not have the conditions or do not use the drugs are identified (specificity).
- If drug and diagnosis misclassification is nondifferential or essentially random, associations between drugs and diagnoses will usually be underestimated (i.e., biased towards the null). If the misclassification is systematic or differential, associations that do not truly exist can appear to be present, or true associations can be overestimated or underestimated.
- Misclassification of drug and diagnosis information obtained from study participants by questionnaires or interviews can be differential or nondifferential and depends on factors such as the training and experience of interviewers, the elapsed time since the events of interest took place, and characteristics of the participants such as their medical status and age.
- Misclassification of drug and diagnosis information obtained from administrative databases is often more likely to be nondifferential than differential, but there is a

greater concern about the accuracy of diagnoses than drugs.

- The medical record is typically used as the gold standard for verifying drug and diagnosis information but it may be incomplete and, with the increasing focus on privacy, such as the US Health Insurance Portability and Accountability Act, may be difficult to obtain.

SUGGESTED FURTHER READINGS

- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–10.
- 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.
- 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.
- Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposures always bias a true effect toward the null value. *Am J Epidemiol* 1990; **132**: 746–8.
- Heerdink ER, Leufkens HG, Herings RM, Ottervanger JP, Stricker BH, Bakker A. NSAIDs associated with increased risk of congestive heart failure in elderly patients taking diuretics. *Arch Intern Med* 1998; **158**: 1108–12.
- Horwitz RI, Feinstein AR. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. *N Engl J Med* 1978; **299**: 1089–94.
- Keeble W, Cobbe SM. Patient recall of medication details in the outpatient clinic. Audit and assessment of the value of printed instructions requesting patients to bring medications to clinic. *Postgrad Med J* 2002; **78**: 479–82.
- Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. In: *Monographs in Epidemiology and Biostatistics*, vol. 10. New York: Oxford University Press, 1986.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.
- Lessler JT, Harris BSH. *Medicaid Data as a Source for Postmarketing Surveillance Information*. Research Triangle Park, NC: Research Triangle Institute, 1984.
- Maclure M, Willett WC. Misinterpretation and misuse of the κ statistic. *Am J Epidemiol* 1987; **126**: 161–9.
- Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.
- Raphael K. Recall bias: a proposal for assessment and control. *Int J Epidemiol* 1987; **16**: 167–70.
- 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.

- Rothman KJ, Greenland S. *Modern Epidemiology*, 2nd edn. Philadelphia, PA: Lippincott, Williams & Wilkins, 1998.
- Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. Cambridge, MA: Cambridge University Press, 2000.
- 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.
- Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol* 1993; **137**: 11251–8.
- 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.

SECTION III

SPECIAL ISSUES IN PHARMACOEPIDEMIOLOGY
METHODOLOGY

Bias and Confounding in Pharmacoepidemiology

The following individuals contributed to editing sections of this chapter:

ILONA CSIZMADI¹ and JEAN-PAUL COLLET²

¹Division of Population Health and Information, Alberta Cancer Board, Calgary, Canada; ²Department of Epidemiology and Biostatistics, McGill University, Montréal, Canada, and Centre for Clinical Epidemiology and Community Studies, SMBD Jewish General Hospital; *Associate Head Research, Department of Pediatrics, UBC Associate Director, Partnership Development, CFRI, Vancouver, BC, Canada.

INTRODUCTION

A major objective of pharmacoepidemiology is to estimate the effects of drugs when they are prescribed after marketing. This is difficult because drug exposure is not a stable phenomenon and may be associated with factors that may also be related to the outcome of interest, such as indication for prescribing. Other examples of factors that must be taken into account include compliance, publicity, and the natural course of the disease. The great challenge of pharmacoepidemiology is thus to obtain an accurate estimate, i.e., “without error,” of the relationship between drug exposure and health status. There are two types of error: *random error* relates to the concepts of precision and reliability, while *systematic error* is related to the concepts of validity and bias. *Accuracy* is the absence of both random and systematic error. Measurement has always

been a key issue in epidemiology; Rothman writes “One way to formulate the objectives of an epidemiologic study is to view the study as an exercise in measurement. The entire process of estimation is in this sense a measurement process” (Rothman and Greenland, 1998).

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGIC RESEARCH

The following example illustrates how pharmacoepidemiology studies may contribute to data required by clinicians, public health administrators, and lawyers. In 1981, Alderslade and Miller presented the results of the National Childhood Encephalopathy Study (NCES), a nationwide case-control study conducted in the United Kingdom by the Committee on Safety of Medicines and the Joint Committee on Vaccination

* Current post

and Immunization. The study was initiated to answer the question of a possible association between diphtheria–tetanus–pertussis (DTP) vaccine and the subsequent development of neurologic disorders. The NCES found that the risk of a severe acute neurologic event was significantly increased within the seven days following DTP vaccine (relative risk (RR) 2.3; 95% confidence interval (CI) 1.4–3.2), and that, one year after the vaccine, 7 of the 241 cases (2.9%) who had died or had a developmental deficit had begun their disease within the seven days following a DTP vaccine compared to only 3 of 478 controls (0.6%), yielding a relative risk of 4.7 (95% CI 1.1–28.0). The results have been used in many court trials by parents of disabled children who were seeking compensation.

Given the importance of the above clinical research, it is not surprising that a heated debate and much publicized controversy erupted when the credibility of the study was compromised by suspicions of bias. Numerous potential biases were identified and implicated in being responsible, wholly or in part, for the results that were observed. These include:

- Referral bias: physicians had been aware of the study objectives and this might have influenced their referral of cases and hence increased the apparent relative risk.
- Information bias: interviewers had not been blinded to study objectives or subjects' clinical status and the date of onset of the neurological disorder was occasionally difficult to ascertain (potentially increasing the apparent relative risk).
- Protopathic bias: possible presence of subclinical neurological disease prior to vaccination, could have falsely increased the relative risk.
- Lack of precision of disease definitions and inclusion of cases thought not to be plausibly related to DTP vaccine (such as Reye's syndrome, hypsarrhythmia, or acute viral encephalopathies), potentially rendered results uninterpretable.

The above example illustrates how study design issues may threaten the validity of results in pharmacoepidemiology research. The study of the sources of bias and the different approaches to preventing bias is thus a fundamental aspect of pharmacoepidemiology. The question of bias in epidemiology has been covered in several good classical epidemiology textbooks, and discussed in some important articles. Pharmacoepidemiology studies, however, may be affected by particular biases more often than other epidemiologic studies. Moreover, the dynamics of bias occurrence over time seems to represent a particularly important phenomenon in pharmacoepidemiology.

In this chapter, we will first describe the most important biases that may affect pharmacoepidemiology studies. We

will then focus on confounding and show that it is sometimes not very easy to separate this threat to study validity from other types of bias (especially selection bias). We will also discuss the dynamics of bias in pharmacoepidemiology and we will show how to deal with the problem of bias and confounding at the design and the analysis stage.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

BIASES IN PHARMACOEPIDEMIOLOGY

In general biases may be grouped into three categories: selection bias, information bias, and confounding. Figure 16.1 shows that *selection bias* is related to the recruitment of study subjects or losses to follow-up, *information bias* is related to the accuracy of the information that is collected on exposure, health status, and covariates that may be confounders or effect modifiers, and *confounding* is related to the pathophysiological mechanisms of disease development, whereby one or several factors acting together can produce an observed effect that may be incorrectly attributed to an exposure of interest.

Selection Bias

Selection bias is a distortion of the measurement of an estimate of effect, which is due to the selection into the study of groups of subjects who differ in characteristics from those in the target population (this bias is also called *sample distortion bias*). In pharmacoepidemiology, four types of selection bias seem particularly important: referral bias, self-selection, prevalence study bias, and the special case of "protopathic bias."

Referral bias can occur if the reasons for referring a patient, to the hospital for instance, are related to the drug exposure status, e.g., when the use of the drug contributes to the diagnostic process. This is a particular problem when an illness presents in a manner such that an accurate diagnosis is not always obtained. For instance, a patient taking nonsteroidal anti-inflammatory drugs and presenting with abdominal pain may be more likely to be suspected as having a gastric ulcer. This patient is therefore more likely to be sent to the hospital for tests for this diagnosis than other patients with similar pain who are not using nonsteroidal anti-inflammatory drugs. A study using patients in the hospital may thus show a strong, but biased, association between mild nonbleeding gastric ulcers and the use of nonsteroidal anti-inflammatory drugs. On the other hand,

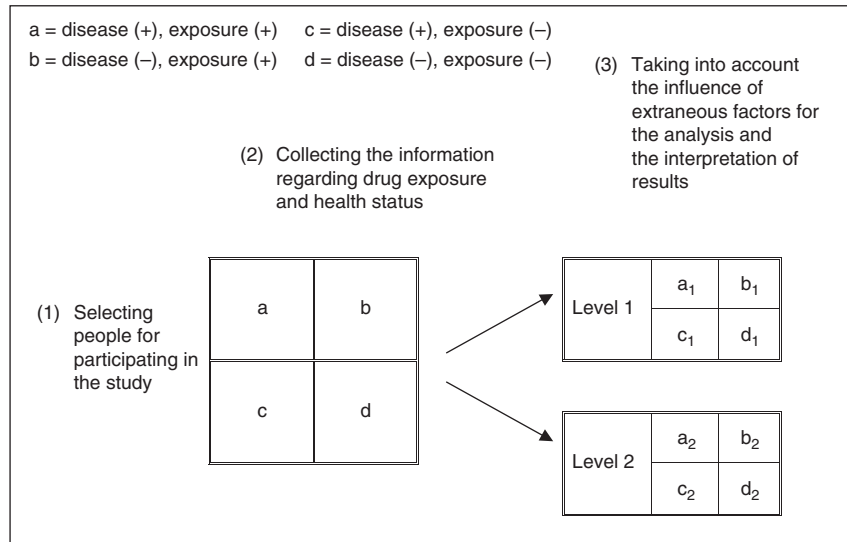


Figure 16.1. Biases in pharmacoepidemiology. (1) *Selection bias* is related to the way people are recruited for the study or are retained during the course of the study. (2) *Information bias* is related to the way the information about study variables is measured during the study. (3) *Confounding* is related to the influence of other variables that are related to both drug use and the outcome, and may be responsible for part or all of the observed effect, the lack of observed effect, or a reversal of the effect.

if one were studying *serious* gastrointestinal bleeding, this might not be a problem.

Similarly, knowledge of a well-established association between deep venous thrombosis and oral contraceptives makes the use of an oral contraceptive a key element in the diagnosis, such that exposed women may be more likely to be subjected to diagnostic tests for venous thrombosis.

Appreciating the potential for referral bias helps in interpreting the results of successive epidemiologic studies conducted at different points in time. An initial study reporting a positive association between drug and disease can begin the referral bias phenomenon whereby over time an increase may occur in the strength of the association, even if the true association remains constant and even if it is actually null.

A general solution to the question of referral bias is to restrict the study to more serious cases of the disease. It can be expected that for most diseases, regardless of previous drug exposure, all serious cases will eventually be diagnosed correctly.

Self-selection bias may occur when study participants themselves decide to participate or to leave a study based on both drug exposure and change in health status. Hence, the association observed in the study sample may not be representative of the real association in the source population. This problem is particularly important in case-control studies or historical cohort studies, because both outcome and exposure

are already manifest when study subjects are recruited. If, for instance, birth defects are studied, we can easily imagine that mothers of affected children who also have “something to report” (i.e., use of medications) may be more (or less) likely to participate (see also Chapter 27). This common problem in pharmacoepidemiology must be controlled at the design level by systematically identifying and recruiting all eligible cases. Losses to follow-up in cohort studies may similarly bias studies if those who drop out belong to a special disease-exposure category. Relying on population-based registries for case and prescription drug ascertainment is an excellent way to minimize the occurrence of selection bias (see Case Example 16.1).

CASE EXAMPLE 16.1:
A STUDY TO DETERMINE THE EFFECT
OF NONSTEROIDAL ANTI-INFLAMMATORY
DRUGS (NSAIDs) ON THE RISK OF
COLORECTAL CANCER

Background

- Evidence from randomized controlled trials and observational study suggests that NSAIDs are protective for colorectal cancer but results have not been consistent.

(Continued)

Issues

- Capturing data that accurately represent NSAID exposure is a challenge since details of patterns of use (e.g., timing, duration, past use of specific formulation, and dose) are not reported accurately by study participants.
- NSAIDs may be prescribed or bought as over-the-counter preparations.
- Numerous potential covariates that may confound the association between NSAIDs and the incidence of colorectal cancer also need to be measured.

Approach

- Use a population-based prescription drug database that provides detailed information pertaining to prescription dose, dispensing date, and formulation.
- Identify colorectal cancer incident cases using population-based cancer registry data; identify population-based controls (sampled to capture person-time at risk).

Results

- Distant past exposure to NSAIDs, >10 years, appears to be protective for colorectal cancer, with the effect becoming more protective with increasing dose.
- More recent exposure does not appear to confer a protective effect.

Strengths

- Incident cases identified using a population-based registry eliminates referral, self-selection, and case prevalence bias.
- Prospectively documented detailed information pertaining to dose and timing of prescribed NSAIDs used to ascertain exposure eliminates problems associated with recall and protopathic bias, minimizes nondifferential and differential misclassification of prescribed NSAID use, and allows for examination risk at various time-windows, including those in the distant past.

Limitations

- Over-the-counter use of NSAIDs was not captured, resulting in a potential for nondifferential misclassification of actual NSAIDs use (although heavy use is most likely prescribed).
- Potential confounding from family history and lifestyle factors (obesity, physical activity, and diet) remain.

Summary Points

- Use of population-based prescription drug databases greatly reduces the various forms of selection bias, which are some of the greatest threats to study validity.
- The possibility and extent to which confounding from covariates is not captured by databases need to be considered in the interpretation of findings from such studies.
- More complicated study designs, e.g., methods using two-stage sampling (discussed later in this chapter), can be considered in order to interview a subsample of study participants who can provide information on important covariates.

Prevalence bias is another type of selection bias that may occur in case-control studies when prevalent cases rather than new cases of a condition are selected for a study. Since prevalence is proportional to both incidence and duration of the disease, an association with prevalence may be related to the duration of the disease rather than to its incidence. An association between drug use and prevalent cases could thus reflect an association with a prognostic factor rather than with incidence. It is possible that a significant association with “good prognosis prevalent cases” might not be confirmed in the whole group of patients defined by incidence. Limiting study recruitment to newly diagnosed or incident cases with a clearly documented calendar time of diagnosis favors the ascertainment of exposures that are relevant to disease incidence.

Protopathic bias is a term that was first used by Feinstein (1985). It may occur “if a particular treatment or exposure was started, stopped, or otherwise changed because of the baseline manifestation caused by a disease or other outcome event.” For instance, people could stop taking aspirin because of the presence of blood in their stools. If the presence of blood were the first expression of colon cancer, we would subsequently find a negative association between current aspirin use and colon cancer. This scenario may occur in pharmacoepidemiology because diseases are often identified late after their first clinical expression and because exposure to drugs may change from day to day, frequently with changes in actual or perceived health status. Such possibilities demonstrate the paramount importance of as full an understanding as possible of the pathophysiologic mechanism of disease development in designing pharmacoepidemiology studies.

Information and Misclassification Bias

Each time participants in a study are classified with regard to their exposure and disease status, there is a possibility of error, i.e. unexposed people may be considered exposed and vice versa. Health status may also be incorrectly classified. This type of error may lead to a *misclassification bias*. It may equally affect case-control and cohort studies. When the error occurs randomly (i.e., independently of the exposure-outcome relationship), it leads to what is referred to as *nondifferential misclassification*. When the degree of error in measuring disease or exposure is influenced by knowledge of the exposure or the outcome status, the misclassification that results is said to be *differential misclassification*. For example, this may occur in cohort studies during the process of data collection, when knowledge of the exposure influences systematically the quality of the information collected about disease outcome. Alternatively, this may occur in case-control studies when knowledge of the disease status influences the quality of the information collected about exposure. This differential misclassification is also called *information bias*. We will describe these two types of misclassification bias in more detail.

Nondifferential or random misclassification occurs when the degree of misclassification is similar for all patients and independent of both exposure and health status (because the instrument is not very reliable, for instance). It may lead to a decrease in the strength of the association between drug and outcome (bias toward the null value). In extreme circumstances, it may even reverse the measure of effect. In pharmacoepidemiology, a number of factors may contribute to the nondifferential misclassification of drug exposure. Figure 16.2 illustrates the different possibilities for misclassifying drug exposure. This is also discussed in more detail in Chapter 15.

Another important misclassification may occur when dichotomizing a patient's exposure as "exposed/not exposed" without taking into account the timing of the exposure. This lack of accuracy in defining exposure may result in an important information bias which may lead to a nonsignificant association overall, whereas in fact, within a specific time-window, there is a very strong association between the drug and the outcome. Similarly, outcome events should also be ascertained within relevant time-windows. Anaphylactic reactions occur rapidly after drug exposure, the risk being very high during a short period of time and null after this initial period. For other outcomes, the risk is likely to decrease with time. For instance, chronic long-term users of a given anti-inflammatory drug are likely to be at a lower risk of gastrointestinal bleeding than new users, because of a survivor effect.

It is also possible that in some circumstances the risk steadily increases with time, due for instance to the cumulative effect of drug exposure (e.g., risk of myocardial toxicity after the use of doxorubicin).

Differential or systematic misclassification can occur when misclassification is related to the exposure-outcome association. Whenever knowledge about disease status in a case-control study or the exposure status in a cohort study influences the validity of the information collected, differential misclassification will be introduced. In pharmacoepidemiology, two situations are commonly responsible for this type of bias: "differential recall" and "differential detection."

- *Recall bias* is an important concern in retrospective studies, e.g., in case-control studies, cases and controls may have a selective memory of their past exposures. In studies of birth defects, for example, mothers with an impaired child may give a more valid and complete report of their exposure to drugs during pregnancy as a result of devoting more time to contemplating the cause of the birth defect. This type of bias may be minimized by selecting controls who are likely to have the same cognitive processes affecting memory of past drug exposures (e.g., alternative birth defects) (see also Chapters 15 and 27).
- *Detection bias* can affect either cohort or case-control studies. It occurs in case-control studies when the procedures for exposure assessment are not similar in cases and controls (e.g., exposure assessment is more thorough among cases). In cohort studies, it occurs when the follow-up procedures for detecting adverse events differ according to the exposure status of the participants. For instance, women taking postmenopausal hormonal supplements are likely to see their doctors more often than other women. These women are therefore more likely to be examined for breast or endometrial cancer, or for the risk of cardiovascular disease. This differential follow-up may lead to an excess number of diagnosed diseases in the treated group and a falsely elevated risk, or to more complete preventive care leading to a decreased risk.

CONFOUNDING

Confounding occurs when the estimate of a measure of association between drug exposure and health status is distorted by the effect of one or several extraneous variables that are also risk factors for the outcome of interest. In the next section, we will present some numerical examples aimed at showing the mechanism of confounding and the statistical

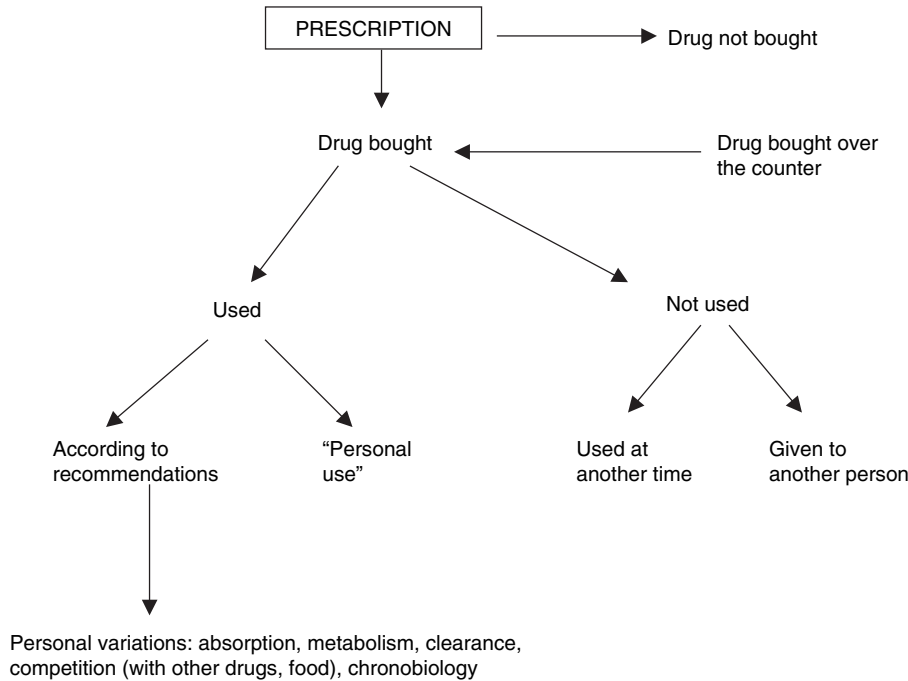


Figure 16.2. Factors influencing drug exposure.

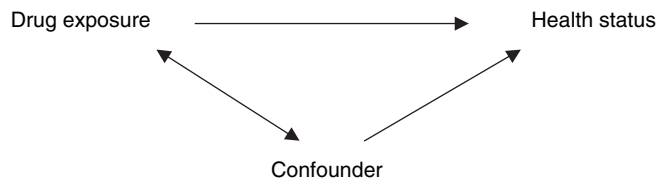


Figure 16.3. Mechanism of confounding.

principles for correcting its effect. We will also describe some important confounders in pharmacoepidemiology and we will show that it is sometimes difficult to distinguish between confounding and selection bias.

Mechanism of Confounding

For a variable to be a confounder, it must be associated with both the drug exposure and the outcome of interest, without being in the causal pathway between the drug exposure and the outcome. In other words, it must represent an independent risk factor. Figure 16.3 shows the classic relationship between the three variables. For example,

when studying the relationship between the use of non-steroidal anti-inflammatory drugs and the occurrence of a gastric ulcer, personal history of gastric problems can be a confounder because: (i) personal history of gastric problems is a risk factor for gastric ulcers, and (ii) a history of gastric problems will modify the probability of physicians' prescribing nonsteroidal anti-inflammatory drugs.

Table 16.1 presents a fictitious example of confounding. A cohort study of deaths associated with the use of drug A was conducted. The comparison group consisted of patients treated with drug B. When data for all patients were considered together, the following results were observed: the risk of death was $202/1100 = 18\%$ among patients

Table 16.1. A cohort study of drug use and risk of death (hypothetical)

Stratum	Treatment	Death		Total	Relative risk
		Yes	No		
All	A	202	898	1100	18%/7% = 2.5
	B	8	102	110	
Severe disease	A	200	800	1000	20%/40% = 0.5
	B	4	6	10	
Benign disease	A	2	98	100	2%/4% = 0.5
	B	4	96	100	

using drug A, and it was $8/110 = 7\%$ among drug B users. The relative risk was therefore $18\%/7\% = 2.5$, indicating a harmful effect of drug A. The study population was then subdivided into two strata: subjects with a severe form of the disease being treated, and those with a more benign form of the disease. When the analysis was conducted within each of these two strata, the direction of the effect was reversed. Among subjects with severe disease, the risks of death were 20% and 40%, respectively, for drugs A and B, for a relative risk of 0.5. The risks of death were lower for subjects with a benign form of the disease, but the relative risk was also 0.5.

These data represent an extreme example of confounding. Obviously, if the estimate of the measure of effect, i.e., the relative risk, is 0.5 among subjects with severe disease and also 0.5 among those with benign disease, the overall estimate of effect, when the data for all subjects are pooled, should also be 0.5. The estimate of 2.5 is not valid, because of confounding by the severity of the disease. An inspection of Table 16.1 shows that subjects with severe disease experience a death rate that is much higher than the rate among those with benign disease. In addition, however, the distribution of study subjects by drug category is not balanced. One thousand subjects with severe disease received drug A. The numbers of subjects in the other subsets of the study population are much smaller. Because of this, the crude estimate of relative risk obtained when the data from the two strata are pooled is heavily influenced by the mortality experience among the 1000 subjects, and this results in the distorted estimate of 2.5.

Table 16.2 gives another example of confounding. In this cohort study, the association between the use of a drug and the risk of allergy was determined. Subjects with and subjects without drug treatment were compared. These data show confounding due to age. Older subjects experienced a high risk of allergy, and a very large group of old subjects without drug treatment was included in the study. Because of this, the overall relative risk was less than the relative risk

Table 16.2. A cohort study of drug use and allergy risk (hypothetical)

Stratum	Treatment	Allergy		Total	Relative risk
		Yes	No		
All	Drug	28	172	200	14%/9% = 1.6
	No drug	102	998	1100	
Young	Drug	8	92	100	8%/2% = 4
	No drug	2	98	100	
Old	Drug	20	80	100	20%/10% = 2
	No drug	100	900	1000	

among young subjects, and also less than the relative risk among old subjects. The confounding was, however, weaker than in the previous example. In the present example, the direction of the association was the same for the overall study population and for the two strata, whereas in the previous example there was a reversal of effect.

Table 16.2 shows another phenomenon. The estimate of relative risk is not the same for young and for old subjects (4 and 2, respectively). This represents *effect modification* or *interaction*: the magnitude of the drug effect on the risk of allergy varies between strata. Statistical tests of heterogeneity (e.g., Breslow-Day or Woolf) are usually used to assess whether such variation between strata can be attributed to random fluctuations, or whether it represents a true effect.

Confounding by Indication for Prescription

The indication for a prescription is probably the most important confounding factor in pharmacoepidemiology since, theoretically, there is always a reason for a prescription and because the reason is often associated with the outcome of interest. It has also been referred to as "indication bias," "channeling," "confounding by severity," or "contraindication bias."

The problem of confounding by indication can be conceptualized in a way that is similar to selection bias, because the decision to prescribe can be viewed as a means of selecting a group of patients. If this selection process is also related to the outcome (which is often the case, especially when considering drug efficacy; see Chapter 21), there is a bias. It is important to note that for a given drug, the nature of confounding by indication is not universal but is directly related to the outcome studied, prescribing practices that may change over time, or from one country to another.

Another characteristic of confounding by indication is the extreme difficulty of controlling for its presence. Although theoretically it is possible to control for it, in practice it

is often impossible to obtain a sufficiently accurate estimate of the effect of this confounder, even when the reason for prescribing seems very straightforward. That is because “indication” is a complex and multifactorial phenomenon involving the physician’s knowledge and many factors which may not be entirely evident and which may act in different directions. Miettinen (1983) provided an example that the preventive use of warfarin can be associated with a 27-fold increase in the risk of thrombotic events, a condition that should actually be prevented by anticoagulation. This paradoxical result was explained by a strong negative confounding effect, i.e., only highly susceptible patients, or those already presenting the first symptoms of thrombosis (see “protopathic bias”) were receiving the therapy.

The example illustrates that it can be very difficult to measure accurately the reasons for prescribing. Consequently, research is needed to elucidate determinants of prescribing practice and behavior. When confounding by indication is suspected, randomized clinical trials are preferred over nonexperimental studies (see Chapter 20). An alternative view of this situation would, on the other hand, consider that, even when confounding by indication is present, the information can be useful for other purposes. For example, it has been postulated that a reported association between β -agonists and asthma mortality was confounded by disease severity. Even in the presence of such a bias, however, a positive correlation between β -agonists and asthma deaths can be informative since the amount of drug use can be a proxy for prognosis. This issue is discussed in much more detail in Chapter 21.

Confounding by Comedication and Other Cofactors

Patients often take more than one drug at a time and it is sometimes difficult to isolate the effect of a specific drug. This question was discussed in the analysis of the Coronary Drug Project (1980), which showed that in the placebo group the risks of death in the 5 years following randomization were 15% and 28.2%, respectively, among compliant patients and noncompliant patients. Beyond a possible selection bias that would relate the better survival to some hypothetical and undetermined factors, the main reason postulated as an explanation for this difference was the fact that patients who were compliant with one drug were also very likely to be compliant with other interventions (e.g., other very effective drugs, diet, physical exercise, etc.). As in the problem of confounding by indication, it is possible to control in part the effect of all other cofactors, but the feasibility of doing so is limited (it is for instance very difficult to quantify compliance precisely), and residual bias is likely to remain.

Confounding and Effect Modification

Confounding and effect modification (or interaction) are both “multivariable phenomena” (i.e. a variable or group of variables which play a role in the observed effect between the drug exposure and the outcome of interest). It is, however, very important to distinguish between the two phenomena, as they have different consequences and require different approaches in data analysis. As we have seen, Tables 16.1 and 16.2 present two hypothetical examples, one with confounding (Table 16.1) and the other one with interaction (Table 16.2). We can see that, as already defined, a third variable (or covariate) is a confounder when it is responsible for part or all of the observed effect. Table 16.1 shows that the “harmful effect of drug A” was in fact due to the selective exposure of patients with the most severe cases of the disease to drug A, while patients with benign disease were more likely to be exposed to drug B. In this situation, the crude estimate of the drug effect represents a combination of two effects: (drug A + severe disease) versus (drug B + benign disease). Table 16.1 also shows that the relative risks for drugs A and B remained constant across the different strata of the confounder. The relative effects of drug A and B were not modified by the severity of the disease. In this situation, severity is said to be a confounder, but not an effect modifier. It is possible to adjust for the difference in distribution of disease severity among the two groups of drug users and obtain an overall relative risk, which in this case will be 0.5, instead of 2.5, the unadjusted crude relative risk.

Table 16.2 shows another example of confounding, but it also shows that the drug effect (measured by the relative risk) varies across the different strata of age. In this situation, age is said to be both a confounder and an effect modifier. It would be possible, like in the previous example, to combine the stratum-specific effects into an overall measure of effect, adjusted for the difference in age distribution. This overall adjusted result, however, is meaningless and may be misleading, as the average result represents a combination of a positive effect for certain patients and negative effects for others. When there is effect modification, the stratum-specific effect provides more information, and is also more interpretable than the single summary figure.

Effect modification corresponds to the statistical concept of interaction. An important point related to interaction is that it is model-dependent: interaction may exist with one parameter measuring the effect, the risk ratio for instance, but not with another parameter, like the risk difference. Moreover, interaction is often a finding at the time of the analysis, as there is generally not enough information for suspecting and quantifying *a priori* its presence.

Table 16.3. Risk of death in asthmatic patients treated by fenoterol and albuterol: change in study results with change in measurement of drug exposure

	Odds ratio	95% Confidence interval
Crude association exposed versus not exposed		
Fenoterol	9.1	3.0–28.1
Albuterol	2.8	1.0–7.6
Results per number of inhalers dispensed		
Fenoterol 0	1.0	Reference group
1–12	4.7	1.1–20.6
13–24	40.5	5.1–319
≥25		17.0–754
Albuterol 0	1.0	Reference group
1–12	3.4	0.9–13.3
13–24	10.0	2.1–46.5
≥25	29.4	5.1–171
Model of continuous exposure		
Fenoterol 100 µg	2.3	1.6–3.4
Albuterol 100 µg	2.4	1.5–3.8

Source: Original data from Spitzer *et al.* (1992).

Effect modification, thus, is often useful for generating new hypotheses and is generally presented as a finding that may be worth studying further. It indicates a variation in the drug effect, according to different levels of a third variable. If this type of finding is confirmed for a particular drug, it may lead to changes in prescribing practice or a better understanding of its mechanism of action. Effect modification therefore warrants thorough examination and careful interpretation.

Effect Modification by Dose or Drug Potency

Different dosage and potency are likely to have different effects that should be presented in the analysis. Reducing the information related to exposure into a dichotomous expression (i.e., exposed versus not exposed) increases the rate of misclassification, biasing the results toward the null. Spitzer *et al.* (1992), for instance, concluded that use of both fenoterol and albuterol was associated with an excess risk of death in asthmatic patients, probably in relation to the severity of the patient's condition. Table 16.3 shows that the dichotomous classification of drug exposure was associated with a three-fold excess risk of death with fenoterol compared to albuterol. However, taking into account (i) the number of inhalers used, and also (ii) the concentration of drug per inhalation (100 µg for albuterol and 200 µg for fenoterol) completely modified the results, such that the risk of death appears similar with the two drugs. This demonstrates the importance in pharmacoepidemiologic

research of considering the possibility of differences in drug dose and drug potency in studying drug effects (see also Chapter 4).

CURRENTLY AVAILABLE SOLUTIONS

HOW CAN ONE DEAL WITH SELECTION BIAS?

Selection bias must be prevented at the design stage, because it cannot be corrected at the analysis stage. The objective is to prevent over- or under-representation of the people who have a particular drug exposure–outcome relationship. This can be achieved in several ways, all of which result in a study population that accurately represents the target population concerning the drug exposure–outcome relationship.

The following strategies can be used:

- Random sampling of the cases and controls (or exposed and non-exposed subjects) to be included in the study from the source population.
- Systematically recruiting a series of consecutive subjects (to prevent self-selection).
- Adopting a well-codified accrual procedure. For example, having a geographic definition of the incident cases minimizes the possibility of introducing referral bias.
- Minimizing the number of subjects lost to follow-up in cohort studies.

- Implementing a tracking procedure for those who drop out of the study, in order to document the reason for dropping out and, if possible, to measure their health status.
- Selecting only incident cases of the condition.
- Assigning random allocation of drug exposure, to prevent self-selection and referral bias.

This “perfect” situation, however, is very difficult to implement, and is generally limited to a small number of people who are followed for a short period of time. Besides ethical, cost, and logistic problems, the experimental design often also creates situations far from real life (see also Chapter 2).

HOW CAN ONE DEAL WITH INFORMATION BIAS?

As with selection bias, the problem of information bias must be resolved at the design stage, since its presence irremediably affects the study validity. Several techniques facilitate an unbiased collection of information:

- *Blinding* (or *masking*) of relevant study personnel is the most important strategy, as it is easier to be neutral when it is not known who is exposed, who is diseased, and what the objectives of the study are. In a cohort study, the data collector should be blinded to the patient’s exposure status and the patient should be unaware of the study objectives. In a case–control study, the data collector should be blinded to the disease status and, if possible, the information related to the past exposure should be collected without knowing the specific objectives of the study.
- *Standardization of the measurement process* for both cases and controls, or exposed and unexposed people, is an essential step when implementing any study. It includes, for instance, the use of standard structured questionnaires, specific training of interviewers, the participation of different observers for different measurements, etc.
- The choices of the criteria for defining drug exposure and disease outcomes are important. Priority should be given to objective, previously defined, standardized criteria.

HOW CAN ONE DEAL WITH CONFOUNDING?

In contrast to information bias and selection bias, it is possible to control the effect of confounding at both the design and the analysis levels. We will present an overview of the different strategies that may be used; we will also describe several approaches that were developed to deal with confounding when working with large although incomplete (no information on confounding) databases, a frequent study design challenge in pharmacoepidemiology.

Dealing with Confounding at the Design Level

There are several ways of controlling for confounding when preparing the design of pharmacoepidemiology studies: randomization, matching, and restriction will successively be described.

If correctly implemented, random allocation of exposure should equalize the distribution of all potentially known and unknown confounders, across the different levels of drug exposure. *Randomization* is thus aimed at making the two groups comparable apart from the assigned exposure (see Case Example 16.2 and further discussion in Chapter 20).

CASE EXAMPLE 16.2: **NUTRITIONAL ANTIOXIDANT** **SUPPLEMENTATION AND RISK OF CANCER**

Background

- Evidence from observational epidemiologic studies has suggested that a high intake of antioxidants such as vitamins C and E and carotenoids through diet or supplements decreases the risk of various cancers (colon, prostate, lung, bladder, and stomach). Randomized controlled trials have not supported these findings.

Issues

- In observational studies individuals self-select themselves to groups with high dietary intakes or supplements of antioxidants.
- People who choose to take supplements generally have other health-promoting lifestyle habits, e.g., they are more physically active, have better overall diets, are leaner, and may smoke less.
- At baseline, in randomized controlled trials covariates are distributed between treatment groups, rendering them comparable on covariates except for the assigned treatment.

Approach

- Randomized controlled trials have been performed in which individual antioxidants and combinations of antioxidants have been studied.

Results

- Randomized controlled trials have not found antioxidants to be protective for cancer.

- Some studies have reported an increase in cancer incidence and mortality with β -carotene supplementation compared with placebo.

Strengths

- Selection bias is avoided at baseline in randomized controlled trials.
- Antioxidant exposure can be classified more accurately due to the assignment of a clearly specified dose.
- Compliance with treatment (supplements) may be more easily determined than in observational designs.

Limitations

- A limited number of dosages can be tested in randomized controlled trials.
- Treatment studied may be very different from the antioxidant exposure in observational studies where dietary intake (exposure of interest) is derived from food sources (e.g., approximately 50 β -carotenes are known to exist in food).
- Although 12 year follow-up studies have been carried out, longer studies may be required for definitive answers to questions about the effect of antioxidants on cancer risk.

Summary Points

- Randomized controlled trials are effective for eliminating selection bias and controlling for known and unknown confounders at baseline.
- Variations in the treatment (dose, formulation) cannot be studied extensively in randomized controlled trials without incurring great expense and much time.
- In randomized controlled trials, replicating the exposure that is studied in observational studies is sometimes extremely difficult.

Matching is another way to control for confounding at the design stage. The objective of matching for both cohort and case-control studies is to make the two compared groups similar with regard to the distribution of selected, known extraneous factors. A matched design requires a matched analysis if the matching variable was truly a confounder. In practice, matching may be difficult, especially when there are several factors on which to match; it may then become costly and time consuming. In case-control studies there may also be a risk of "overmatching," which increases the similarities between groups. In such a case, because of the matching there is artifactually a reduction in the true

differences between cases and controls in the levels of the exposure of interest.

Restriction of the design to only one level of the confounding factor is the simplest, but also the most reductive, way of dealing with confounding. For instance, studying the effect of a drug in only one category of age will protect against the occurrence of confounding by age. The generalizability of the study, however, may be confined to this age group.

Dealing with Confounding at the Analysis Level

Standardization

Standardization represents a classical method of controlling for confounding; it is often used for comparing vital statistics from populations that have different age or sex distributions, especially in occupational epidemiology. A standardized (or summary) rate is thus a weighted average of stratum-specific rates.

Standardized rate = $\sum_i W_i I_i / \sum_i W_i$ where W_i and I_i represent, respectively, the stratum-specific weight and the stratum-specific incidence rate. The ratio of standardized rates provides an estimate of the effect of the exposure, where the ratio of standardized rates = $\sum_i W_i (a_i / N_{1i}) / \sum_i W_i (b_i / N_{0i})$ and N_{1i} and N_{0i} represent, respectively, the stratum-specific size of the populations which are exposed (N_{1i}) and not exposed (N_{0i}) to the drug, and a_i and b_i represent the number of cases in each stratum which are, respectively, exposed (a_i) and not exposed (b_i) to the drug.

There are two methods of standardization: direct and indirect. If, for example, we wish to control for age, in direct standardization the age stratum-specific event rates of the groups to be compared are applied to a standard population. The expected number of events are calculated and summed across the age categories and then divided by the size of the standard population to provide age-adjusted rates for each group. These unconfounded rates can then be compared. In indirect standardization, stratum-specific populations of the groups to be compared are multiplied by age-specific rates of a particular event from a standard population to provide the expected number of events in each group. The expected numbers of events for each group (had they had the age-specific rates of the standard population) are then divided by the observed number of events to provide standardized ratios, which can be compared between the groups. The number of factors that can be managed by standardization is limited (i.e., two or three). Pharmacoepidemiologic research usually requires the manipulation of more than three factors and often, therefore, requires the use of other techniques like mathematical modeling.

Stratification

Stratification is another way of obtaining a summary rate ratio that adjusts for confounding. It is performed in two stages: the first stage requires the computation of a stratum-specific rate ratio for each level of the stratifying (confounding) variable. The second stage involves pooling the results into a single estimate that represents the “overall” effect of the drug, adjusted for the effect of the confounding factor. Standardization and stratification both pursue the same objective of obtaining a pooled estimate of the drug effect.

There are several ways of pooling the stratum-specific measures of effect into an overall estimate. A classic approach consists of defining the weights as proportional to the inverse of each stratum’s variance (i.e., weighting the contribution of each stratum by its statistical stability). The most popular approach to this has been proposed by Mantel and Haenszel (1959) for the odds ratio in case-control studies, and provides a formula which is easy to compute:

$$\text{Adjusted rate ratio} = [\sum_i W_i (a_i/N_{1i})(b_i/N_{0i})] / \sum_i W_i$$

Mantel–Haenszel odds ratio = $\sum_i a_i (N_{0i}/T_i) / \sum_i b_i (N_{1i}/T_i)$ where T_i , N_{0i} and N_{1i} represent, respectively, the stratum sample size, the number of unexposed controls, and the number of exposed controls; a_i and b_i represent the number of cases in, respectively, the exposed and the unexposed groups.

Stratification is performed in order to have very little or no variation of the confounder within each stratum of the confounding variable. Stratification, therefore (like “adjustment,” described below), requires an accurate measurement of the confounding variable to fulfill its objective. Nondifferential misclassification of a confounder may lead to the persistence of some confounding.

The limitation of stratified analysis is that, each time a new factor is added, stratum-specific cell sizes become smaller, and the probability of having people not exposed or not sick in each stratum becomes larger. The stratum-specific estimate of the measure of association cannot then be computed, and the stratum does not provide any statistical information. In this situation, it is better to use a multivariate approach, modeling the relationship between exposure to all factors of interest and the outcome.

Multivariate Analysis and Modeling

Determining the relationship between risk factors and outcomes using a mathematical model allows the assessment of many factors at the same time. According to the prespecified model of relationship (this choice is a crucial one), a parameter of effect will be estimated for each risk

factor. This estimate represents the individual contribution of the factor for the risk of the outcome, adjusted for all other factors in the model. However, in order to maintain reasonably stable estimates of parameters there are general rules of thumb for ensuring an optimal sample size: approximately 10 observations are required per factor in a multiple regression model where the outcome is continuous, and approximately 10 events per factor for logistic regression models where the outcome is binary. Hence, while multivariate analysis provides a more efficient means of controlling for several factors simultaneously, the adequacy of the data in meeting these requirements must be examined. Most of the important models are derived from *general linear* equations and are very powerful tools that require sophisticated skills in biostatistics and epidemiology, which are beyond the scope of this chapter but are well dealt with in books on multivariate statistical methods.

Dealing with Confounding when Working with Large Drug Databases

Standardization, stratification, and modeling all require an accurate measurement of the confounding variable. An increasing number of pharmacoepidemiology studies are now performed using large databases of previously collected data (see Chapters 11 and 12). One of the major limitations of these studies is the impossibility of adjusting for potential confounders not included in the database. An example would be a study of the relationship between low-dose oral contraceptives (OC) and the risk of myocardial infarction. Databases are likely to provide accurate information for OC prescription and the occurrence of myocardial infarction, for example, but are very unlikely to provide any information on smoking habits, a strong risk factor for myocardial infarction, which is also related to OC use. Not being able to control for such an important confounder precludes the valid study of this association, because the results will obviously be biased if OC users are also heavy smokers.

In order to address this type of study design problem White (1982) and Walker (1982) suggested sampling a fraction of the study population to gather information about confounding variables. This information can then be used in the analysis to obtain covariable-adjusted estimates of the parameters of interest. This approach, referred to as “two-stage sampling,” was further developed by Cain and Breslow (1988) for multivariate analysis. Efficiency is the essence of this approach, motivated by the desire to use resources optimally. In this approach, stage 1 represents the study population, for example the cases and controls in a case-control study. Individuals for stage 2 are selected

according to their disease–exposure characteristics. The balanced design is often more efficient than random and disease- or exposure-based sampling; it consists in having an equal number of individuals in each cell of the second stage 2×2 table. This strategy decreases the occurrence of small cells (responsible for large variance) by forcing an overrepresentation of individuals who belong to small groups in the exposure–disease cross-classification. The sampling fractions that lead to the second stage sample are typically different for each exposure–disease category, creating a selection bias that must be corrected in the analysis. The two-stage design permits the detection of, and adjustment for, confounding. Interaction can also be evaluated. Schaebel *et al.* (1997) proposed software for sample size estimation for two-stage sampling.

Propensity Scores for Efficient Adjustment

The propensity score, proposed and developed by Rosenbaum and Rubin (1983), is an innovative statistical approach that can be used to increase the comparability of treatment groups in the absence of randomized treatment assignment. Defined as the conditional probability of being treated, given an individual's covariates, the objective of propensity score analysis is to simulate randomized controlled trial treatment groups in order to estimate a causal treatment effect. One should caution, however, that this objective is achievable only to the extent that all covariates related to treatment assignment have been well measured. Unmeasured relevant covariates will remain potential sources of bias in the estimation of effect. Recently, propensity scores have been used in clinical, health services, and pharmacoepidemiologic research.

The propensity score is estimated using logistic regression where the exposure or treatment of interest is the dependent variable and covariates related to treatment assignment are the independent variables. The probability or propensity of receiving a treatment, given the observed covariates, is then determined for each subject. Individuals with similar propensity scores are comparable since they are similar in their propensity to receive the treatment under study. The effect of treatment can be estimated using this propensity score (i) as a matching variable prior to analysis, (ii) during analysis to define quintiles for stratification, or (iii) as a covariate in regression analysis. By using the propensity score in the analysis, the effects of all of the prognostic covariates used in estimating it are removed from the estimation of the treatment effect, thus reducing bias. For any given value of the propensity score the unbiased average treatment effect is estimated for that propensity score if the

treatment is “strongly ignorable,” given the covariates. In other words, the probability of being assigned a treatment is dependent only on the propensity score covariates and otherwise treatment assignment is random.

Propensity score analysis is not intended to replace well-designed randomized controlled trials (RCTs). Rather, it is a viable and appropriate tool that should be considered in clinical and pharmacoepidemiologic research when RCTs are not feasible due to ethical or cost concerns or when outcomes are extremely rare. It is also an analysis that can be conducted to generate hypotheses for future clinical trials or to provide sufficient evidence to establish clinical equipoise in order to justify an RCT.

Sensitivity Analysis

In this chapter we have discussed the various sources of bias and confounding that can affect pharmacoepidemiology studies. While threats to validity need to be minimized at the design stage of a study, potential biases that have not been accounted for need to be addressed in the presentation of results and discussion.

Traditionally, the “uncertainty” about study findings has only been addressed in “Discussion” sections of research reports and publications. The limited scope and impact of this practice have recently been highlighted in epidemiology literature. Sensitivity analysis, defined as a quantitative analysis of the potential for systematic error, is a more formal approach used to communicate this uncertainty with respect to the validity of findings. There are several advantages to using this quantitative approach compared with limiting the discussion to a qualitative assessment when presenting results. Sensitivity analysis not only brings to the forefront the important issues related to the validity of results but it also provides a means for presenting objective evidence that may be used by readers to evaluate the magnitude of the threat to validity. In addition, the results of a sensitivity analysis can provide direction for, or suggest, areas of future research. As articulated by Greenland (1998), “it [sensitivity analysis] can be viewed as an attempt to bridge the gap between conventional statistics, which are based on implausible randomization and random-error assumptions, and the more informed but informal inferences that recognize the importance of biases, but do not attempt to estimate their magnitude.”

THE FUTURE

One great challenge facing the future in pharmacoepidemiology is the ability to control adequately for “indication

for prescribing” at the analysis stage. This requires obtaining a valid and complete ascertainment of the reasons for prescribing drugs. This could be accomplished by adopting very strict, standardized, and measurable criteria for prescribing drugs. Whenever it is not possible to clearly differentiate the respective effects of the drug and the underlying medical conditions, implementing randomized clinical trials in the postmarketing phase should be considered (see Chapter 20).

It is also interesting to consider bias and confounding in pharmacoepidemiology under a dynamic perspective. Confounding is susceptible to change with time as well, as the decision to prescribe depends directly on the physician. We could even conceptualize the ultimate objective of research in medicine and pharmacoepidemiology so as to bias the prescribing of physicians, as only patients who may benefit from the drug should receive it. With this perspective, identification of a change in efficacy or variability in drug effect according to a patient’s characteristics or other factors (i.e., the presence of effect modification) should subsequently induce a change in physicians’ prescribing to take into account this information. A recent example of this is the change in hormone replacement therapy (HRT) prescribing since the release of the Women’s Health Initiative results in July 2001. Any further studies of the effect of HRT might then be biased by an altered indication for prescribing if it is not taken into account. We may therefore view confounding by indication as a natural and positive consequence of integrating the results of research into medicine and even an objective of pharmacoepidemiology rather than simply a nuisance while estimating the real effect of a drug.

Another important issue facing pharmacoepidemiology is the ability to measure drug exposure accurately (see also Chapter 15). This should be accomplished by more accurately defining exposure time-windows, by better considering dosage and potency, and by accurately measuring drug use; the study of compliance thus represents a highly promising domain of research with regard to the study of drug effects (see Chapter 25). The development of population pharmacokinetics and pharmacodynamics (see Chapter 4), as well as pharmacogenetics (see Chapter 18), should also provide useful information for interpreting pharmacoepidemiologic results with regard to drug exposure. Finally, propensity scores opened a new way for efficient adjustment. This approach could in the future allow adjusting for reasons of prescribing drugs, but only to the degree that indication can be measured. In the end, since it is usually not possible in any study to be certain that confounding has been completely controlled for, sensitivity analyses

should be routinely incorporated into the discussion section of study of results.

Key Points

- A thorough understanding of the potential sources of bias and confounding in research studies is a prerequisite to designing studies that minimize their influence on results.
- Pharmacoepidemiology studies are susceptible to being affected by the threats to validity that are common to all epidemiology studies.
- Confounding by indication is an example of a potential threat to validity that is unique to pharmacoepidemiology and requires special consideration, particularly in the design of nonexperimental studies.
- The existence of numerous population-based health and prescription drug databases has facilitated the design of observational studies where some aspects of selection bias, information bias, and confounding are more easily controlled.
- With a distinctive set of challenges and unique set of resources, pharmacoepidemiologists are developing and implementing innovative study designs and new approaches to analyses in order to ensure the validity of results from nonexperimental studies.
- Due to the complexity of pharmacoepidemiology studies, a final examination of the validity of study results in a formal sensitivity analysis will facilitate better informed interpretation and discussion of results.

SUGGESTED FURTHER READINGS

- Bjelakovic G, Nikolova R, Simonetti RG, Gluud C. Antioxidant supplements for prevention of gastrointestinal cancers: a systematic review and meta-analysis. *The Lancet* 2004; **364**: 1219–28.
- Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. *Am J Epidemiol* 1988; **28**: 1198–206.
- 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.
- Collet JP, Sharpe C, Belzile E, Boivin JF, Hanley J, Abenhaim L. Colorectal cancer prevention by non-steroidal anti-inflammatory drugs: effects of dosage and timing. *Br J Cancer* 1999; **81**: 62–8.
- Coronary Drug project research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* 1980; **303**: 1038–41.

- 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.
- Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research*. Philadelphia, PA: WB Saunders, 1985.
- Greenland S. Basic methods for sensitivity analysis and external adjustment. In: *Modern Epidemiology*, 2nd edn. Philadelphia, PA: Lippincott-Raven, 1998; pp. 343–57.
- Guess HA. Behavior of the exposure odds ratio in a case-control study when the hazard function is not constant over time. *J Clin Epidemiol* 1989; **42**: 1179–84.
- Hanley JA, Csizmadia I, Collet JP. Two-stage case-control studies: precision of parameter estimates and considerations in selecting sample size. *Am J Epidemiol* 2005; **162**: 1225–34.
- Lash TL, Silliman RA. A sensitivity analysis to separate bias due to confounding from bias due to predicting misclassification by a variable that does both. *Epidemiology* 2000; **11**: 544–9.
- Maclure M, Schneeweiss S. Causation of bias: the episcopes. *Epidemiology* 2001; **12**: 114–22.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **22**: 719–48.
- Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983; **2**: 267–71.
- Miettinen OS. Confounding and effect modification. *Am J Epidemiol* 1984; **100**: 350–3.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.
- Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984; **79**: 516–24.
- Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia, PA: Lippincott-Raven, 1998.
- Schaubel D, Hanley J, Collet JP, Sharpe C, Boivin JF. Controlling confounding when studying large pharmacoepidemiologic databases: the two-stage sampling design. *Am J Epidemiol* 1997; **146**: 450–8.
- 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.
- Wacholder S, Carroll RJ, Pee D, Gail MH. The partial questionnaire design for case-control studies. *Stat Med* 1994; **13**: 623–4.
- Walker AM. Anamorphic analysis: sampling and estimation for covariate effects of both exposure and disease are known. *Biometrics* 1982; **38**: 1025–32.
- White JE. A two-stage design for the study of the relationship between a rare exposure and a rare disease. *Am J Epidemiol* 1982; **115**: 119–28.

Determining Causation from Case Reports

Edited by:

JUDITH K. JONES

The Degge Group, Ltd. and Adjunct Faculty, Georgetown University, George Washington University School of Public Health, Washington, DC, USA.

INTRODUCTION

An important component in the evaluation of reports of suspected adverse drug reactions in the clinical setting, or in a clinical trial, is the judgment about the degree to which any reported event is causally associated with the suspected drug. In reality, a particular event is either caused or not caused by a particular drug, but the current state of information almost never allows a definitive determination of this dichotomy. A number of approaches to determination of the probability of a causal drug–event association have evolved over the past several years. This chapter will briefly review several of the current approaches and uses and their origins. It will then review the regulatory context on this topic and the applications of these methods.

THE CLINICAL PROBLEM ADDRESSED BY METHODS OF DETERMINATION OF THE LIKELIHOOD OF CAUSALITY

The basic clinical problem is that a clinical event occurs within the milieu of a number of possible causal factors and either occurs independently or in some way its occur-

rence is partially or totally linked to one or more of the potential causative agents. The primary task is to conduct a differential diagnosis and 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 approach to evaluating causality in chronic disease epidemiology. However, in the latter case, assessment relates to events in one or more defined population studies. In individual case reports of suspected adverse reactions to a medicinal product submitted to the manufacturer, regulatory agency, or the literature, data are often very incomplete and causal assessment is very challenging. Because these reports often represent a clinician's suspicion of a causal association, both the reporter and the evaluator of these cases often make an implicit judgment of causality. Single reports often have several attributes that represent obstacles to causality assessments, specifically:

1. The usual focus of assessment is a clinical event *suspected* by the reporter of being associated with exposure to a drug or other medicinal product. This can often bias the collection of data required to evaluate other possible causes.

2. The data on exposure to the suspect and other concomitant drugs are often incomplete, usually missing precise information on duration, actual dose ingested, and past history with the exposures.
3. The data available on the adverse event, including its onset, characteristics, and time course, are also typically incomplete, in part 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. Complete 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 medicinal product cause, rather than a differential diagnosis.

Adverse reactions to drugs 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. One challenge has 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 reintroducing 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 determination and the impact of that inference on any actions taken. If the determination 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, the assessment might logically be less rigorous. Conversely, if, for example, continuation of a clinical trial or drug development program depends 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, e.g., the US Food and Drug Administration (FDA)'s regulations for reporting in clinical trials of adverse events that are "reasonably" associated with a drug (CFR 21 : 312.22), the need for consistent and reliable methods of causality determination has become more important.

HISTORICAL PERSPECTIVES: DEVELOPMENT OF CONCEPTS OF CAUSALITY FOR ADVERSE REACTIONS

Thinking about the causality of adverse reactions has evolved at approximately the same time in two disciplines:

(i) in epidemiology, and (ii) in the evaluation of individual case reports of suspected adverse reactions.

In the 1950s, Yerushalmy and Palmer, drawing upon the Bradford Hill epidemiologic criteria as well as the Koch-Henle postulates for establishing causation for infectious diseases, developed a set of proposed criteria for causality in epidemiologic studies. These evolved into five criteria for the causal nature of an association: the consistency, the strength, the specificity, the temporal relationship of the association, and the coherence, or biological plausibility, of the association. These criteria continue to be generally used in chronic disease epidemiology. In the adverse drug reactions field, the typical approach to case reports was to consider the events as possibly associated with the drug simply if there were a number of similar reports. Considerations of pharmacologic plausibility, dose-response, and timing factors were sometimes implicit, but seldom explicit. This unstructured approach by what has been termed "global introspection" is still used in some cases and judgments expressed in terms of a *qualitative* probability scale, for example "definite," "probable," "possible," "doubtful," or "unrelated."

The recognized subjective nature of this global reasoning led Nelson Irey (1976), a pathologist at the US Armed Forces Institute of Pathology, and Karch and Lasagna (1977), two clinical pharmacologists, to develop approaches that used very similar basic data elements 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 ("de-challenge");
4. the result of reintroducing the drug ("re-challenge");
5. other data supporting an association, e.g., previous cases.

The object for these causality assessments was either a single case or a group of cases from an ill-defined exposed population, and there was only a vague resemblance to the Bradford Hill criteria. Thus, in either a single report or even a series of cases, there would be no way to evaluate the consistency, strength, or specificity of the association (with some exceptions). The temporal relationship does apply in both approaches, and in some cases the coherence or biological plausibility has been included in the criteria for single cases when a specific pharmacologic mechanism or dose response is present.

Following the introduction of these new methods for the assessment of suspected adverse drug reactions, a large number of other approaches were developed, in the form of

Table 17.1. A summary of the information categories in the method of Kramer *et al.* (1979) for determining causality of adverse drug reactions

Axis ^a	Information category	Number of questions in the axis ^b	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

^aAxis 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 = definite to less than 0 = unlikely. ^bEach 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.

algorithms, decision tables, and in one case as a diagrammatic method. Most of these methods, reviewed and summarized in two conference monographs and in a textbook chapter, share the basic elements originally suggested by Irey (1976) and Karch and Lasagna (1977). A summary of some of the elements is presented in Table 17.1, and they are categorized and described further in the next section of this chapter.

To address the inherent limitations of all these methods, a more advanced approach to causality assessment, based on Bayes probability theorem, subsequently emerged. 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 known background information and all details of the case (see Case Example 17.1).

ACTUAL AND POTENTIAL USES OF CAUSALITY ASSESSMENT

There are certain settings where causality assessments are currently conducted and there are other potential uses of these assessments.

Pharmaceutical Manufacturers

Pharmaceutical sponsors must view causality assessment for events associated with their drugs from the standpoints of both regulatory requirements in different countries as well as product liability. In the US, 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” (US Code of Federal Regulations 21:314.80). In the FDA Investigational New Drug (IND) regulations there is an implicit requirement for causality assessment to determine the need for reporting certain types of events in clinical trials, based on the requirement for the reporting of serious unexpected events where there is a “reasonable possibility” that the events may have been caused by the drug (CFR 21:312.22). A disclaimer is noted that such a report of a serious unexpected event does not constitute an admission that the drug caused the events. These IND regulations do not provide criteria or a suggested method.

Outside of the US, many regulatory agencies have requested or implied some type of evaluation to minimize the number of nonspecific events reported, described below.

Drug Regulators

Drug regulators’ internal use of causality assessment of spontaneously reported postmarketing adverse reactions has varied considerably. Most countries’ drug regulators have some method of approaching causality and this method has been formally considered in a European Community Directive resulting in a general consensus on the causality terms used by the European Union member states. Further, since 1994, a formal method of causality assessment for reports of vaccine-associated adverse events has been instituted by Health Canada.

In the US, a simple algorithm was used in the early 1980s, based on only the criteria of timing, de- and re-challenge, and confounding factors. It specifically excluded consideration of previous literature reports based on the reasoning

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, i.e., the causality assessment was a mechanism for identifying the best documented cases—those with a “probable” or “highly probable” association—and not for actual causality assessment. The publicly available files from the FDA’s Adverse Event Reporting System database consistently carry the caveat “a cause and effect relationship has not been established.”

Publishers of Reports of Adverse Reactions

The medical literature containing case reports of suspected adverse reactions has largely avoided the issue of causality, with the exception that The *Annals of Pharmacotherapy* now requires that the Naranjo method 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. Further, many reports do not provide information on confounding drug therapy or medical conditions, data elements considered to be essential for considering causality. Harambaru *et al.* (1990) compared 500 published reports with 500 spontaneous reports with respect to the availability of information needed in most standard causality assessments. Although the published reports contained significantly more information, there were very sparse data on both alternate causes/other diseases and other drugs in *both* types of reports. This failure to make a structured approach to case reports was recognized in the early 1980s and publications from a consensus conference proposed that the publication of case reports requires at minimum the five elements of the criteria for causality (e.g., details of timing, the nature of the reaction, discontinuation and re-introduction, and alternate causes based on prior history). A similar proposal for publication guidelines is forthcoming in 2006 from a task force of the International Society of Pharmacoepidemiology.

Other Applications

There are other settings where standard assessments of causality could be useful. These include evaluation of serious events in clinical trials in drug development, in formal consensus evaluation of new serious postmarketing spontaneous reports by both sponsors and regulators, in the clinical setting where suspected adverse reaction should be a common component of the differential diagnosis, and possibly in the courtroom and even the newsroom.

METHODOLOGIC PROBLEMS TO BE ADDRESSED IN ASSESSMENT OF CAUSALITY

The problem to be solved in determining whether an event is caused by a drug 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 intrinsic nature of drug-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 oculo-mucocutaneous syndrome caused by practolol). In addition, since drugs are used to treat illnesses, drug-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 grudgingly 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. 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. This latter view has spurred the evolution of causal evaluation from expert consensual opinion based on global introspection to structured algorithms, and to elaborate Bayesian probabilistic approaches.

CURRENTLY AVAILABLE SOLUTIONS

Four basic types will be described, chosen as illustrative examples and because they have been widely described in various publications.

UNSTRUCTURED CLINICAL JUDGMENT/GLOBAL INTROSPECTION

The most common approach to causality assessment is probably unstructured clinical judgment. One or more experts

are 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.

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, which is precisely the task of causality assessment. 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.

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 a numerical scale. Despite these concerns, global introspection for evaluation of adverse events continues to be used. For example, the WHO International Centre for Drug Monitoring in Uppsala, Sweden that collects the spontaneous reports from national centers worldwide has published causality criteria ranging from “certain” to “unassessable/unclassifiable” that essentially represent six levels of global introspection, though they generally incorporate consideration of the more standard criteria for causality.

ALGORITHM/CRITERIAL METHOD WITH VERBAL JUDGMENTS

These methods range from simple flow charts posing 10 or fewer questions to lengthy questionnaires containing up to 84 items. However, they share a common basic structure essentially based on the five elements of timing, re-challenge, de-challenge, confounding, and prior history of the reaction. Information relevant to each element is elicited by a series of questions, the answers to which are restricted to “yes/no” (and, for some methods, “don’t know”). This type of approach is used by some drug regulatory agencies, such as in Australia, and has been used previously in an FDA algorithm.

When compared to global introspection, there is a great improvement in the consistency of ratings among reviewers. Since the consideration of each case is segmented into its components (e.g., timing, confounding diseases, etc.), 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. 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 one week represented a sufficient duration of drug exposure to be consistent with a drug–event association.

ALGORITHMS REQUIRING SCORING OF INDIVIDUAL JUDGMENTS

Several 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 quantitative methods have found applications in a number of settings, ranging from evaluations of suspected adverse reactions by hospital committees to applications by some regulatory authorities, as in France. They are also used, although sometimes only in a research context, by some pharmaceutical manufacturers. One of the more practical examples of this type was that discussed previously, by Naranjo *et al.* (1990). This algorithm is shown in Figure 17.1. A more recent version called RUCAM has six criteria with three or four levels of scoring for each criterion to derive an overall score. This has recently been applied in evaluation of adverse events in HIV clinical trials, and was more recently cited by Lee (2000) in review of methods of assessment of hepatic injury.

PROBABILISTIC METHODS

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 17.2. 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;
2. what the likelihoods are, or are not, for drug causation of *each* of the components of the specific case, including history, timing, characteristics, de-challenge and its timing components, re-challenge, and any other factors, such as multiple re-challenges.

CAUSALITY ASSESSMENT NARANJO SCORED ALGORITHM				
QUESTION	ANSWER			SCORE
	Yes	No	Unk	
Previous reports?	+1	0	0	_____
Event after drug?	+2	-1	0	_____
Event abate on drug removal?	+1	0	0	_____
+ Re-challenge?	+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 17.1. A critical scored algorithm illustrated by the method of Naranjo *et al.* (1990) in wide use. This particular method uses some of the basic data elements as well as more details of the history and characteristics of the case, and a score is designated for the response to each question.

POSTERIOR ODDS = PRIOR ODDS X LIKELIHOOD RATIO

$$\frac{P(D \rightarrow e) \mid B, C}{P(D \nrightarrow E) \mid B, C}$$

↓

Overall probability

$$\frac{P(D \rightarrow E) \mid B}{P(D \nrightarrow E) \mid B}$$

↓

Epidemiology and clinical trial data

$$\frac{P(C \mid D \rightarrow E)}{P(C \mid D \nrightarrow E)}$$

↓

Individual case data (history, timing, case character, de-challenge, etc.)

P Probability

D → E Drug caused event

D ↛ E Drug did not cause event

B Baseline information

C Case c event

Figure 17.2. 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. 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.

The full application of this method requires detailed knowledge of the clinical event, its epidemiology, and relatively specific information about the event’s characteristics and kinetics over time. Examples have been published for several types of events, including Stevens–Johnson syndrome, renal toxicity, lithium dermatitis, ampicillin-associated colitis, agranulocytosis, and Guillain–Barré syndrome. 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 cohort studies, when standard methods of statistical analysis will not provide sufficient clues as to causality because of inadequate sample size.

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 sparse data on the incidence of most events and their occurrence (1) in the presence and (2) in the 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. Although this lack of information is a current limitation, it represents both an important challenge and a framework for structuring further understanding. For this reason, 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 ten-fold difference in the estimate of the timing does not materially modify the overall posterior odds estimate, further efforts to determine the “best” estimate would not be worthwhile.
3. Because of the multi-step 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 complementary method developed by Begaud and colleagues using the Poisson method for estimating the probability of rare events in populations, has been published recently by Zapater *et al.* (2002). 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. An elegant and detailed evaluation of six representative algorithmic methods, carried out by Pere *et al.* (1986), identified standard evaluation criteria and evaluated 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.

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 and use of quantitative or the Bayesian method would be more appropriate. 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 Jones, 1984).
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. In some cases this results from data collection being focused to a much greater degree, as has been initiated in France, where special reporting forms based on disease-specific criteria for events are being developed.
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.

CASE EXAMPLE 17.1: **GENERAL APPROACH TO CAUSALITY** **ASSESSMENT OF A SUSPECTED ADVERSE** **DRUG REACTION**

Background

- Application of a structured causality assessment can serve as a template for standardized assessment of the different elements of a case report of an adverse event that is suspected to be caused by a drug.

(Continued)

- A formal approach helps to avoid “global introspection,” which is naturally biased by an assessor’s own knowledge and preconceived notions.

Issue

- A 68-year-old Caucasian female with a history of severe osteoarthritis and occasional angina pectoris is given a prescription for a moderately potent nonsteroidal anti-inflammatory drug (NSAID) to be used at the lowest dose. Five days after starting therapy, she experiences two episodes of angina on exertion in the same day. Upon contacting her physician, she is instructed to discontinue the NSAID. She experiences no further anginal attacks, but is also distressed at the return of her painful arthritis. On her own, she restarts the NSAID. Within 3 days, she experiences four episodes of angina on exertion and notifies her physician, who instructs her to stop the NSAID and come in the next day for an examination. She has no further episodes of angina after discontinuing the drug.

Approach

- Identify key components that may help to determine causality and a differential diagnosis: (1) timing (in both cycles); (2) other causes of the event and prior history of the event; (3) de-challenge (improvement upon discontinuing suspected drug); (4) re-challenge (return of event on reintroduction of drug); and (5) prior history of the drug–event association (not always used if event is the first observed).

Results

- Timing is consistent in both cycles with initiation of drug intake, but also with the ability to increase exercise level (see other factors, below).
- Other factors that may explain the event include prior history of angina, an intermittent condition, and increased exercise with relief of the arthritis.
- De-challenge is positive both times.
- Re-challenge is positive.
- There are no data on dose or blood levels (biological plausibility).
- Global introspection on this case might reflect the assessor’s information biases. One oriented to NSAID gastrointestinal effects might consider this a masked gastrointestinal event, another might consider the exercise tolerance as a cause, while another might suspect a direct vasoconstrictor effect of the NSAID. Despite the

presence of positive de-challenges and a re-challenge, if the algorithm used gives “other factors” higher weight to offset the score or weight of de-challenge/re-challenge, then many methods would find this event “possibly related” to the NSAID.

- Confounding by the increased exercise tolerance and the intrinsic intermittency of the angina would make a stronger association less likely in the absence of further data.
- A Bayesian approach would try to identify, from clinical trials or observational data on the pattern of angina pectoris, the prior odds of the pattern of the two separate intermittent episodes of angina episodes occurring in (1) the absence or (2) the presence of the NSAID. Then the actual case is analyzed by determining the likelihood (if drug caused or not) of *each* of the following components: history of angina (e.g., the ratio of the probability of the pattern of angina given NSAID causation / the probability of the angina pattern given non-NSAID causation); timing of the episodes relative to drug start; de-challenge and its timing in both episodes; and re-challenge and its timing.

Strengths

- Use of specific categories of data to understand the dynamics of the event and the dosing and possible pharmacologic properties of the drug directs the deliberation to consider more possibilities to explain the pathophysiology and ultimately the probability of drug causation of the event.
- The Bayesian approach is more structured and intensive.

Limitations

- Methods that are verbal or have numerical scores that do not always relate to actual pathophysiology may lead to misleading conclusions.
- Methods that require more data are often frustrated by the simple lack of data needed, and, when applying the Bayesian method specifically, the lack of baseline data in the population to make prior probability or likelihood judgments.

Summary Points

- Applications of a structured causality method can standardize and help to reduce biases in assessing the possible cause–effect relationship of an event to a particular drug exposure.

- The components of causality assessment methods can help to structure data collection on individual and groups of cases; ultimately, these aggregate data can improve the description of the event of interest, and possibly its relationship to a drug, or the disease of indication.
- The detailed probabilistic and explicit approach in the Bayesian method can, if data are available, provide a basis for developing more precise statements of the hypothesis that is posed in a spontaneous report of a suspected adverse drug reaction.

THE FUTURE

The field of adverse reaction causality assessment described in the preceding sections has many unresolved issues, both methodological and practical. Although there was an original hope that there would be some basis for a consensus method, the current state of the field would suggest that this is not likely to be the case. 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. 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 not been the case. Third, the misuse of judgment terms or scores within the legal arena has generated concern, particularly given the fact that there is not a reliable standard terminology.

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. Further work in this area can have a major impact on:
 - a. the collection of better information on the different drug-associated events, using data collection instruments tailored to the event of interest, and
 - b. the better definition of the dynamics and, ultimately, the pathophysiology and mechanisms of certain types of drug-induced conditions.
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, some methods are now being automated, including the French method. Convenient access to the proper questions, arrayed in logical order, as well as background data meeting quality criteria on the state of information to date, has the potential for considerably improving 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, “root cause analysis” is used to identify the important contributors to adverse events in clinical settings. This approach creates functional maps of possible contributing factors to not only identify a cause but also determine methods of preventing it. Another approach is the *N-of-1* trial, which can evaluate the causality of adverse events in individuals, particularly those who have experienced multiple reactions to drugs.

In conclusion, the topic of causality of adverse reactions continues to represent a challenge. With increased awareness 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.

Key Points

- Applications of a structured causality method can standardize and help to reduce biases in assessing the possible cause–effect relationship of an event to a particular drug exposure.
- The use of a clinical non-structured approach (“global introspection”) to assess adverse events believed to be associated with a drug has been shown to yield inconsistent results between raters, and its lack of structure does not further the development of the hypothesis raised in the report of the event.
- The choice of a method is based upon the use of the judgment; if pivotal to continued development of a drug, the most rigorous methods such as the Bayesian approach may help; if used to sort out well-documented cases that may probably be associated with a drug, then simple algorithms or scoring algorithms will usually suffice.
- The components of causality assessment methods can help to structure data collection on individual and groups of cases; ultimately, these aggregate data can improve the description of the event of interest, and possibly its relationship to a drug, or the disease of indication.
- The detailed probabilistic and explicit approach in the Bayesian method can, if data are available, provide a basis for developing more precise statements of the hypothesis that is posed in a spontaneous report of a suspected adverse drug reaction.

SUGGESTED FURTHER READINGS

- de Finetti B. *Theory of Probability*. New York: John Wiley & Sons, 1974.
- 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.
- 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.
- Herman RL, ed. *Drug–Event Associations: Perspectives, Methods, and Users*. Proceedings of Drug Information Association Workshop, Oct 30–Nov 2, 1983, Arlington, Virginia. *Drug Inf J* 1984; **18**: 195–352.
- Hutchinson TA, Lane DA. Assessing methods for causality assessment. *J Clin Epidemiol* 1989; **42**: 5–16.
- Investigational New Drug Procedures, 21 CFR § 312.22, 1987.
- Irey NS. Adverse drug reactions and death: a review of 827 cases. *JAMA* 1976; **236**: 575–8.
- Jones JK. Uses of drug–event assessment. *Drug Inf J* 1984; **18**: 233–40.
- Jones JK. Determining causation from case reports. In Strom BL, ed., *Pharmacoepidemiology*, 4th edn. Chichester: John Wiley & Sons, 2005; pp. 557–70.
- Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. *Clin Pharmacol Ther* 1977; **21**: 247–54.
- 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.
- 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.
- Lee WM. Assessing causality in drug-induced liver injury. *J Hepatol* 2000; **33**: 1003–15.
- Louik C, Lacouture PG, Mitchell AA, Kauffman R, Lovejoy FH, Yaffe SJ, Shapiro S. A study of adverse reaction algorithms in a drug surveillance program. *Clin Pharmacol Ther* 1985; **38**: 183–7.
- Naranjo CA, Lanctot KL, Lane DA. The Bayesian differential diagnosis of neutropenia and antiarrhythmic agents. *J Clin Pharmacol* 1990; **30**: 1120–7.
- Pere JC, Begaud B, Haramburu F, Albin H. Computerized comparison of six adverse drug reaction assessment procedures. *Clin Pharmacol Ther* 1986; **40**: 451–61.
- Venulet J, Berneker GC, Ciucci AG, eds. *Assessing Causes of Adverse Drug Reactions*. London: Academic Press, 1982.
- Zapater P, Such J, Perez-Mateo M, Horga JF. A new Poisson and Bayesian-based method to assign risk and causality in patients with suspected hepatic adverse drug reactions. *Drug Saf* 2002; **25**: 735–50.

Molecular Pharmacoepidemiology

Edited by:

STEPHEN E. KIMMEL

University of Pennsylvania School of Medicine, 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 (2003) 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 25), and, last but not least, the genetic profile of the patient.

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 pharmacoepidemiology 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 methodological 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. 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, 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%) of variability in drug disposition and medication

effects. 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.

DEFINITIONS AND CONCEPTS

Genetic Variability

Building on the success of the various human genome initiatives, it is now estimated that there are approximately 30 000 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). We have also learned that there is a great deal of inter-individual 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 >1% of the population. Each person has inherited two copies of each gene (one from the paternal chromosome and one from the maternal chromosome). The term allele refers to the specific nucleotide sequence 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 2 common missense (i.e., amino acid changing) variants per gene. 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. 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. These regions define genomic block structure 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, Pharmacogenomics, and Molecular Pharmacoepidemiology

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*, along with including studies of genetic variability on drug response, also encompasses approaches simultaneously considering data about thousands of genotypes in drug discovery and development, as well as responses in gene expression to existing medications. Although the term “pharmacogenetics” is sometimes used synonymously with pharmacogenomics, the former usually refers to a candidate-gene approach as opposed to a genome-wide approach in pharmacogenomics (both discussed later in this chapter).

Molecular pharmacoepidemiology focuses on the effects of genetics on clinical outcomes from medication use. This is unlike pharmacogenetic and pharmacogenomic studies, which typically are designed to examine intermediate endpoints between drugs and outcomes (such as drug levels, pharmacodynamic properties, or surrogate markers of drug effects). Molecular pharmacoepidemiology answers questions related to:

1. the population prevalence of SNPs and other genetic variants;
2. evaluating how these SNPs alter disease outcomes;
3. assessing the impact of gene–drug and gene–gene interactions on disease risk;
4. evaluating the usefulness and impact of genetic tests in populations exposed, or to be exposed, to drugs.

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 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.

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. 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. Metabolism of medications can either inactivate their effect or convert an inactive prodrug into a therapeutically active compound. 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. Thus, patients using CYP2D6-dependent antipsychotic drugs (e.g., haloperidol) who are poor metabolizers (low CYP2D6 activity) are more than four times more likely to need antiparkinsonian medication to treat side effects of the antipsychotic drugs than extensive metabolizers. The decreased metabolic activity of CYP2D6 may also lead to lower drug efficacy, as illustrated for codeine, which is a prodrug that is metabolized to the active metabolite, morphine, by CYP2D6.

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), which has polymorphisms that have been associated with, for example, resistance to antiepileptic drugs. It has been found that patients with drug-resistant epilepsy are more likely to have the CC polymorphism of ABCB1, which is associated with increased expression of this transporter drug-efflux protein. Of note, and consistent with the complexities of molecular pharmacoepidemiologic research noted later, the ABCB1 polymorphism fell 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, presumably 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. 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 (see Case Example 18.1).

CASE EXAMPLE 18.1: PHARMACODYNAMIC EFFECT OF GENETIC VARIANTS

Background

- *Regular use* of inhaled β -agonists for asthma may produce adverse effects and be no more effective than *as-needed* use of these drugs.

Question

- Can genetic polymorphisms in the β_2 -agonist receptor alter responsiveness to inhaled β -agonists?

(Continued)

Approach

- Perform genetic analysis within a randomized clinical trial of regular versus as-needed use of inhaled β -agonists.
- Compare the effects of multiple genetic polymorphisms on drug response.

Results

- Regular use of inhaled β -agonist is associated with decline in efficacy among those with β_2 -AR-16 mutations but not among those with other mutations tested.
- There is no effect of genotype in those using inhaled β -agonists in an as-needed manner.

Strengths

- Randomized trial design eliminates confounding by indication for frequency of medication use.
- Candidate genes enhances biological plausibility.

Limitations

- Multiple polymorphisms tested on multiple outcomes lead to concern of false positives.
- Linkage disequilibrium: polymorphisms identified could be “innocent bystanders” by being linked to the true causative mutations.

Summary Points

- Genetic polymorphisms of drug targets may alter drug response.
- Because of the concern of false positives and/or linkage disequilibrium, replication studies and mechanistic studies remain critical to identifying true putative mutations that alter drug response.
- Effects of a gene may vary by the pattern of drug use, making it important to consider all aspects of drug use (dose, duration, frequency, regularity, etc.) in molecular pharmacoepidemiology studies.

Pharmacodynamic gene–drug interactions may also affect the risk of adverse reactions. For example, a polymorphism in the gene coding for the bradykinin β_2 receptor has been associated with an increased risk of angiotensin converting enzyme (ACE) inhibitor-induced cough.

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, antihypertensive medications that work by a particular mechanism, such as increasing sodium excretion, may have different effects depending on the susceptibility of the patient to the effects of the drug. Patients with a polymorphism in the α -adducin gene have greater sensitivity to changes in sodium balance. A case–control study has suggested that those with the α -adducin polymorphism may be more likely to benefit from diuretic therapy than those without the polymorphism.

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 twenty Caucasians, is considered an important genetic risk factor for deep vein thrombosis and embolism. 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 recently also been linked to the differential thrombotic risk associated with third-generation oral contraceptives compared with second-generation oral contraceptives. Despite this strong association, Vandenbroucke *et al.* (1996) 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. Therefore, they came to the conclusion 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. This highlights another important role of molecular pharmacoepidemiology: determining the utility and cost-effectiveness of genetic screening to guide drug therapy.

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. Depending on the combination of these genotypes, the individual might have a different response in terms of both efficacy and toxicity (see Table 18.1). In the simplified example in Table 18.1, there is one genetic variant that alters drug metabolism and one genetic variant that

Table 18.1. Hypothetical response to medications by genetic variants in metabolism and receptor genes

Gene affecting metabolism ^a	Gene affecting receptor response ^a	Drug response	
		Efficacy (%)	Toxicity (%)
Wild-type	Wild-type	70	2
Variant	Wild-type	85	20
Wild-type	Variant	20	2
Variant	Variant	35	20

^a Wild-type associated with normal metabolism or receptor response and variants associated with reduced metabolism or receptor response. Source: Modified from Evans and McLeod (2003).

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.

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. The enzyme primarily responsible for the metabolism of warfarin to its inactive form is the cytochrome P450 2C9 variant (CYP2C9). Case Example 18.2 illustrates both the logical progression of pharmacogenetics through molecular pharmacoepidemiology and the complexity of moving pharmacogenetic data into practice.

CASE EXAMPLE 18.2: THE COMPLEXITY OF THE PROGRESSION AND APPLICATION OF MOLECULAR PHARMACOEPIDEMIOLOGY RESEARCH

Background

- Warfarin is a narrow therapeutic index drug. Underdosing or overdosing, even to a minimal degree, can lead to significant morbidity (bleeding and/or thromboembolism).

Question

- Can genetic variants be identified and used to alter the dosing of warfarin and thus improve safety and effectiveness?

Approach and Results

- Multiple study designs have been used to address this question.
- First, pharmacogenetic studies identified the effect of CYP2C9 polymorphisms on warfarin metabolism.
- Second, a case–control study comparing warfarin patients requiring low doses versus patients not requiring low doses found that CYP2C9 polymorphisms were associated with lower dose. By design, this study selected subjects based on warfarin dose requirements, not genotype, and could only determine that lower doses of warfarin were more common among those with CYP2C9 variants. The other associations noted were between lower dose requirements and bleeding, not between genotype and bleeding.
- Third, in order to address the clinically relevant question of bleeding, a retrospective cohort study was performed that demonstrated an increased risk of bleeding among patients with at least one CYP2C9 variant. The retrospective nature of the study left unanswered the question of whether knowing that a patient carries a variant can alter therapy in a way that can reduce risk.
- The recent development of an algorithm 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. However, this has not yet been tested prospectively.

(Continued)

Strength

- A logical series of studies, each with its own strengths and limitations, has improved our understanding of genetic variability in response to warfarin.

Limitations

- No study has yet determined whether one can reduce adverse events and enhance the effectiveness of warfarin by knowing a patient's genetic make-up.
- Along with CYP2C9, numerous other polymorphisms have been identified that may alter response to warfarin, raising the possibility that CYP2C9 may, by itself, not be sufficient to maximize safety and effectiveness.

Summary Points

- The process of fully understanding the effects of polymorphisms on drug response requires multiple studies and often substantial resources.
- Our understanding of genetic variants has progressed so rapidly that new questions are often raised that have implications for clinical practice even as old questions are answered.
- Before using genetic data to alter drug prescribing, prospective evaluation is needed.

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 methodological problems of pharmacoepidemiology must be addressed in molecular pharmacoepidemiology. These problems include those of chance and power, confounding, bias, and generalizability (see Chapters 2 and 3).

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 genes, and the possibility that a presumptive gene–drug response relationship may be confounded by the racial and

ethnic mixture of the population studied. Thus, the methodological 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 pharmacoepidemiology studies rely on proper identification of putative genes. In addition, in all research of this type, usage 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 methodological challenges of studying interactions, minimizing type I and type II errors, and accounting for confounding, particularly by population admixture (defined below).

GENE DISCOVERY: GENOME-WIDE VERSUS CANDIDATE GENE APPROACHES

There are two primary, but not mutually exclusive, approaches for gene discovery: candidate gene association studies and genome-wide scans. 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. Each approach has strengths and limitations. Candidate gene studies have the advantage of using knowledge of molecular biology, biochemistry, and physiology to elucidate biologically plausible associations of genotypes with outcomes of interest. However, it remains a challenge to choose appropriate candidate genes because the relevant biological information needed to choose candidates may not be available. A major challenge facing studies that measure the statistical relationship of a clinical outcome with candidate disease genes is to characterize the functional significance of genetic variants. In contrast, genome-wide approaches avoid the need for biological plausibility in identifying genes for study, and instead use large-scale genetic or genomic information to search for genes with effects on phenotypes of interest. These approaches use the wealth of genomic information to scan the genome for important genes. However, the identification of these genes leads directly back to the need for biological information that explains the causal mechanism for the gene's association effect. Ultimately, genome-wide scans will identify genes that will become candidate genes, and will thus also require knowledge about gene function. A major limitation of genome-wide approaches is the limited ability

to assess gene and variant function based on nucleotide sequence information alone. This is most likely to be true when variants do not alter an amino acid or disrupt a well-characterized motif that affects protein function or structure. It is likely that only a small subset of variants will actually confer small to moderate effects on phenotypes that are causally related to disease risk.

INTERACTIONS

Along with examining the direct effect of genes and other biomarkers on outcomes, molecular pharmacoepidemiology studies must often be designed to examine effect modification between medication use and the genes or biomarkers of interest. That is, the primary measure of interest is often the role of biomarker information on the effect of a medication. For purposes of simplicity, this discussion will use genetic variability as the measure of interest.

Effect modification is present if there is a difference in the effect of the medication depending on the presence or absence of the genetic variant. This difference can be either on the multiplicative or additive scale. On the multiplicative scale, interaction is present if the effect of the combination of the genotype and medication exposure relative to neither is greater than the product of the measure of effect of each (genotype alone or medication alone) relative to neither. On the additive scale, interaction is present if the effect of the combination of the genotype and medication exposure is greater than the sum of the measures of effect of each alone, again all relative to neither.

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. 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 18.2). The second is to present a 2×4 table (also shown in Table 18.2). In the first example (stratified analyses), 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 minus medication; with medication minus 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. In the example given in Table 18.2, multiplicative interaction would be assessed by comparing the odds ratio for the combination of genotype and medication exposure to the product of the odds ratios for medication alone and genotype alone. Multiplicative interaction would be considered present if the odds ratio for the combination of medication and genotype (A in Table 18.2) was greater than the product of the odds

Table 18.2. Two ways to present effect modification in molecular pharmacoepidemiologic studies using case-control study as a model

Genotype	Medication	Cases	Controls	Odds ratio	Information provided
<i>Stratified analysis</i>					
+	+	a	b	ad/bc	Effect of medication versus no medication among those with the genotype
	−	c	d		
−	+	e	f	eh/fg	Effect of medication versus no medication among those without the genotype
	−	g	h		
<i>2 × 4 Table</i>					
+	+	a	b	ah/bg = A	Joint genotype and medication versus neither
+	−	c	d	ch/dg = B	Genotype alone versus neither
−	+	e	f	eh/fg = C	Medication alone versus neither
−	−	g	h	Reference	Reference group

ratios for either alone ($B \times C$). 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, presenting the data in both manners is probably 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 that 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 3). It is easy to appreciate the potential for type I error in a molecular pharmacoepidemiology study that examines, simultaneously, the effects of multiple genetic factors, the effects of multiple nongenetic factors, and the interaction between and among these factors. One of the reasons cited for nonreplication of study findings in molecular pharmacoepidemiology is type I error. Limiting the number of associations examined to those of specific candidate genetic variants that are suspected of being associated with the outcome is a "standard" method to limit type I error in pharmacoepidemiology. However, with increasing emphasis in molecular pharmacoepidemiology studies on identifying all variants within a gene and examining multiple interactions, this method of limiting type I error may not be desirable. 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, the ability to detect a gene–response relationship is likely to require relatively large sample sizes to avoid type II error (concluding that there is no association when in fact one does exist). The sample size requirements for studies that examine the direct

effect of genes on medication response will be the same as the requirements for examining direct effects of individual risk factors on outcomes. With relatively low prevalences of polymorphisms and often low incidence of outcomes (particularly in studies of adverse drug reactions), large sample sizes are typically required to detect even modest associations. For such studies, the case–control design has become a particularly favored approach for molecular pharmacoepidemiology 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. 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. There were a total of 1038 participants in the study. When comparing the effect of diuretic use with no use and comparing the effect of the genetic variant with the nonvariant allele, all 1038 participants were available for comparison (Table 18.3). However, when examining the effect of diuretic therapy versus nonuse among those with the genetic variant, only 385 participants contributed to the analyses. Of note, this study presented the data for interaction in the two ways presented in Table 18.2.

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. Population admixture is simply a manifestation of confounding by ethnicity, which can occur if both baseline disease risks and genotype frequency vary across ethnicity. The larger the number of ethnicities involved in an admixed population, the less likely that population stratification can be the explanation for an observation. Empirical data show that carefully matched, moderate-sized case–control samples in Caucasian populations are unlikely to contain levels of population admixture that would result in significantly inflated numbers of false-positive associations. There is the potential for population structure to exist in African American populations, but this structure can be eliminated by removing recent African or

Table 18.3. Gene–exposure interaction analysis in a case–control study

Diuretic use	Adducin variant	Cases	Controls	Odds ratio (OR) for stroke or myocardial infarction
0	0	A ₀₀ 103	B ₀₀ 248	1.0
0	1	A ₀₁ 85	B ₀₁ 131	1.56
1	0	A ₁₀ 94	B ₁₀ 208	1.09
1	1	A ₁₁ 41	B ₁₁ 128	0.77

Case control OR for diuretic use in variant carriers: $OR_{\text{variant}} = A_{11}B_{01}/A_{01}B_{11} = 41 \times 131/85 \times 128 = 0.49$

Case control OR for diuretic use in wild-type carriers: $OR_{\text{wild-type}} = A_{10}B_{00}/A_{00}B_{10} = 94 \times 248/103 \times 208 = 1.09$

Synergy index = $OR_{\text{variant}}/OR_{\text{wild-type}} = 0.45$

Case-only OR = $A_{11}A_{00}/A_{10}A_{01} = 41 \times 103/94 \times 85 = 0.53$

Source: Adapted from Psaty *et al.* (2002).

Caribbean immigrants, and limiting study samples to resident African Americans. Based on the current literature that evaluates the effects of confounding by ethnicity overall, and specifically in African Americans, there is little evidence that population stratification is a likely explanation for bias in point estimates or incorrect inferences. Nonetheless, population admixture must be considered in designing and analyzing molecular pharmacoepidemiology studies to ensure that adequate adjustment can be made for this potential confounder.

CURRENTLY AVAILABLE SOLUTIONS

GENE DISCOVERY: GENOME-WIDE VERSUS CANDIDATE GENE APPROACHES

As discussed in the “Methodologic problems” section, there are two primary approaches for gene discovery: candidate gene association studies and genome-wide screens. The latter approach relies on linkage disequilibrium (LD), defined above as the correlation between alleles at two loci. The genome-wide association approach uses DNA sequence variation (e.g., SNPs) found throughout the genome, and does not rely on *a priori* functional knowledge of gene function. Therefore, this approach can be used to identify new candidate genes or regions. These approaches rely 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 epidemiological 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. 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. It has been hypothesized that studies that consider LD involving multiple SNPs in a genomic region (i.e., a haplotype) can increase the power to detect associations by 15–50% compared with analyses involving only individual SNPs. 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 should 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. On the one hand, investigators can begin by studying simple haplotypes and then move to more complete haplotype characterization. Similarly, investigators can study candidate genes and then move to the characterization of haplotypes in the candidate region to better understand the spectrum of genetic variability that may be associated with the trait of interest. Ultimately, both individual SNPs with known functional significance as well as haplotype data may be required to fully understand the basis of human pharmacogenetics. Haplotypes will capture information about the total genetic variability in a genomic region, but knowledge about etiologically significant functional genetic changes will still be required to understand the biological basis of phenotypically relevant associations.

INTERACTIONS

Along with traditional case–control and cohort studies, the case-only study can be used for molecular pharmacoepidemiology studies designed to examine interactions between genes and medications (see Case Example 18.3).

CASE EXAMPLE 18.3: THE CASE-ONLY STUDYBackground

- Identifying interactions between genes and medications on outcomes often requires large sample sizes. In addition, proper selection of a control group in case-control studies can be challenging.

Question

- Can a case-only study be used to more efficiently identify interactions between medications and genes?

Approach

- Select cases with the outcome of interest.
- Measure the association between genetic variants and the medication of interest among the cases.

Results

- 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.
- 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 18.3 footnote).

Strengths

- Eliminates the need to identify controls, which is often a major methodological and logistical challenge in case-control studies.
- Can result in greater precision in estimating interactions compared with case-control analyses.
- 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.

Limitations

- Relies on the assumption that the use of the medication is unrelated to the genotype. 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.

- Does not allow assessment of the independent effects of medication use or genotype on outcome.
- Interaction can only be interpreted on a multiplicative scale.

Summary Points

- Case-only studies can be used to measure the interaction between genetic variants and medications.
- Case-only studies eliminate the difficulty and inefficiency of including a control group.
- Case-only studies rely on the assumption of independence between medication use and genetic variants among those without disease, an assumption that may not be met.
- Case-only studies can be used within the context of a case-control study or using large-scale databases.

TYPE I ERROR AND REPLICATION

Given concerns of type I error (along with other methodological 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.* (2003) 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.).

Lohmueller *et al.* (2003) 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 was seen, while only 5% were expected under the null hypothesis. This replication was not explainable by publication bias or false positives due to ethnic stratification. The explanation for the lack of replication among the other associations could have been due partly, but not wholly, to different LD or other population patterns or population-specific modifiers (genes

and/or environments). Another possibility is the “winner’s curse” phenomenon, which predicts that the initial positive report overestimates the “true” value. An additional consequence of this phenomenon is that replication studies may require larger sample sizes since the actual replication effects 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 3). One approach to increasing the sample size of molecular pharmacoepidemiology studies is to perform large, multicenter collaborative studies. Another is to assemble large, relatively homogenous populations for multiple studies, such as the deCode project (<http://www.decode.com/nrg1/markers/>). This project has established a cohort of almost 100 000 inhabitants of Iceland with available genetic data that can be used for both population- and genome-wide linkage studies. This represents half of the total adult population of Iceland and

includes more than 90% of people over the age of 65. This project has also generated debate about the process of informed consent.

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 24).

POPULATION ADMIXTURE

As presented above, although population stratification is unlikely to be a significant source of bias in epidemiological association studies, a number of analytical approaches exist to either circumvent problems imposed by population genetic structure, or that use this structure in gene identification. 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 and learn about population structure. It further uses the data derived from these unlinked markers to adjust the association test statistic. Using a similar conceptual model, the “genomic control” discussed above in “Type I Error and Replication” uses a similar conceptual model. If unlinked markers are associated with disease, one possible explanation is the presence of population admixture.

THE FUTURE

Ultimately, 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 will 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, in selection of a specific therapeutic class of drug to treat a disease, and in avoidance of using specific medications in those at high risk for adverse drug reactions. Somewhat relatedly, the measurement of genotypes in randomized trials of medications is a powerful way to assess the effect of genetic variation on medication response in a relatively unbiased fashion.

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 22). Veenstra (2004) has developed a set of criteria for evaluating the potential clinical and economic benefits of pharmacogenetic testing. 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 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.

Just as for all research, the ethical, legal, and social implications of genetic testing must be considered and addressed (see also Chapter 19). 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. There also is a concern that medicines will be developed only for the most common, commercially attractive, genotypes, leading to “orphan genotypes.” Cases

of idiosyncratic side effects in susceptible patients based on genotyping also point to the question of whether it would be feasible to develop alternative drugs targeted to these nonresponders (e.g., carriers of HLA B5701 in the example of abacavir). For sure these would be “orphan drugs,” according to current epidemiologic data on HIV and associated drug-induced side effects in the US and Europe. Would it then be economically realistic for the pharmaceutical industry to invest in such drugs? Subsetting of patients according to response to therapy based on gene–drug interactions may have a plethora of consequences for the pharmaceutical marketplace; some will be predictable and others will not. Although beyond the scope of discussion for this chapter, it is important to note that ethical, legal, and social implications of genetic testing in general are critically important issues that will need to be addressed along with the methodological challenges of molecular pharmacoepidemiologic research.

Finally, it is important to recognize that genetic and other biomarkers are unlikely to fully explain the variability in drug response. Continued efforts at characterizing other causes of drug failures and adverse reactions and understanding how these causes interact with genetic polymorphisms are critical to fully characterize the variability of drug response and to ultimately improve patient outcomes.

Key Points

- Genes can affect a drug response via: alteration of drug pharmacokinetics, pharmacodynamic effects on drug targets, and gene–drug interactions in the causal pathway of disease.
- Molecular pharmacoepidemiology is the study of the manner in which molecular biomarkers (often, but not exclusively, genes) alter the clinical effects of medications in populations.
- Molecular pharmacoepidemiology answers questions related to: the population prevalence of SNPs and other genetic variants; evaluating how these SNPs alter disease outcomes; assessing the impact of gene–drug and gene–gene interactions on disease risk; and evaluating the usefulness and impact of genetic tests in populations exposed, or to be exposed, to drugs.
- Identifying genes that alter drug response for molecular pharmacoepidemiology studies can be done using a candidate gene or genome-wide approach; these approaches are really complementary, not mutually exclusive.
- The methodological challenges of molecular pharmacoepidemiology are closely related to issues of statistical interactions, type I and type II errors, and confounding.
- Case-only studies can be used to measure the interaction between genetic variants and medications, thus eliminating the difficulty and inefficiency of including a control group. However, they rely on the assumption of independence between medication use and genetic variants among those without disease, an assumption that may not be met.
- Given concerns of type I error (along with other methodological concerns such as uncontrolled confounding, publication bias, and linkage disequilibrium), a key issue in molecular epidemiology is the ability to replicate association study findings.
- Because genetic variability leading to phenotypic expression of complex diseases results from the relatively small effects of many relatively low prevalence genetic variants, the ability to detect a gene–response relationship is likely to require relatively large sample sizes to avoid type II error. Methods to ensure adequate sample sizes include: the use of large, multicenter collaborative studies; assembly and genotyping of large, relatively homogenous populations for multiple studies; and meta-analysis.

- Population stratification can distort the gene–medication response association. Although unlikely to be a significant source of bias in well-controlled epidemiological association studies, a number of analytical approaches exist to either circumvent problems imposed by population genetic structure, or that use this structure in gene identification.
- The ability of genes and other biomarkers to improve patient care and outcomes needs to be tested in properly controlled studies, including randomized controlled trials. Similarly, the cost-effectiveness of such approaches must be justifiable given the additional costs of genetic testing in clinical care.
- The ethical, legal, and social implications of genetic testing must be considered and addressed, just as they must be considered for all research.

SUGGESTED FURTHER READINGS

- Aithal GP, Day CP, Kesteven PJ, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet* 1999; **353**: 717–19.
- 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.
- Evans WE, McLeod LJ. Pharmacogenomics—drug disposition, drug targets, and side effects. *N Engl J Med* 2003; **348**: 528–49.
- Gage BF, Eby C, Milligan PE, Banet GA, Duncan JR, McLeod HL. Use of pharmacogenetics and clinical factors to predict the maintenance dose of warfarin. *Thromb Haemost* 2004; **91**: 87–94.
- 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.
- 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.
- 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.
- 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.
- Phillips KA, Veenstra DL, Oren E, Lee JK, Sadec W. Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. *JAMA* 2001; **286**: 2270–9.
- 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.
- Reidenberg MM. Evolving ways that drug therapy is individualized. *Clin Pharmacol Ther* 2003; **74**: 197–202.
- 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.
- 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.
- 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.
- 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.
- Vesell ES. Pharmacogenetics: multiple interactions between genes and environment as determinants of drug response. *Am J Med* 1979; **66**: 183–7.
- 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.
- Williams-Jones B, Corrigan OP. Rhetoric and hype: where’s the “ethics” in pharmacogenomics? *Am J Pharmacogenomics* 2003; **3**: 375–83.

Bioethical Issues in Pharmacoepidemiologic Research

The following individuals contributed to editing sections of this chapter:

KEVIN HAYNES,¹ JASON KARLAWISH,^{1,2} and ELIZABETH B. ANDREWS³

¹Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; ²University of Pennsylvania, Department of Medicine, Division of Geriatrics, Philadelphia, Pennsylvania, USA; ³Research Triangle Institute Health Solutions, Research Triangle Park, North Carolina, USA.

INTRODUCTION

The discipline of research ethics has assumed a largely protectionist posture, largely because of a series of unfortunate scandals and the public outcry that ensued. 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. Themes that run through many of the underlying historical scandals are scientists' failure to adequately review and disclose research risks and potential benefits, and their failure to obtain explicit permission from research subjects. As a result of these events, review by an institutional review board (IRB) and full informed consent have become the cornerstones of the protection of human subjects from research risks.

These and other requirements have been remarkably effective in defining the limits of ethical research and 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 patient-oriented interventional research are applied to other kinds of research. In particular, standard protections in patient-oriented interventional research are not always easily exported and applied to the very different challenges of observational epidemiologic research.

The human subjects involved in observational pharmacoepidemiologic research are quite different than the human subjects in randomized clinical trials for which traditional ethics guidelines are built. The "human subjects" in pharmacoepidemiology studies are often data points in a data set, obscuring the human subject component. 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 epidemiology research are not the usual health risks of research that can be balanced against the potential health benefits of research. Harm to the subject due to an experimental intervention is not the issue in pharmacoepidemiologic research. The chief risk is the violation of confidentiality. While investigators and ethics review boards

may be able to balance medical 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 ethical structure to the growing field of epidemiology and pharmacoepidemiology. These guidelines have addressed four broad categories of ethical issues in epidemiology research: obligations to society, obligations to funders and employers, obligations to colleagues, and obligations to subjects. The most challenging guideline has clearly proven to be the investigators' obligation to protect subjects. 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 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. Investigators face considerable challenge in protecting patient privacy and confidentiality in a way that accomplishes research goals accurately and efficiently. This challenge lies at the heart of the ethics of pharmacoepidemiologic research.

The goal of this chapter is to present an overview of the balance and specifically the challenges that arise when the principles of research ethics are applied to issues surrounding privacy and confidentiality. The 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. The chapter next will discuss three strategies to balance the need for scientific rigor with the need to respect ethical requirements and the challenges investigators face in applying them: delinking subject identifiers from their information, modifications to bioethics review board review, and modifications to subject informed consent requirements. The 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 PHARMACOEPIDEMOLOGIC RESEARCH

RESEARCH

The summary statement of the Belmont Report from the US National Commission for the Protection of Human

Subjects (1979) defined "research" as any activity designed to "develop or contribute to generalizable knowledge." Unfortunately, this definition creates a challenge for pharmacoepidemiologic researchers and ethics review boards, because it is not always easy to characterize the intent of the person who generates the knowledge. 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. It is not clear, given the definition based on "generalizable knowledge," whether the program should be construed as research, clinical care, or even as a quality improvement activity. 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.

The paradigmatic *practice* of epidemiology is public health case finding and in pharmacoepidemiology specifically is the surveillance of adverse drug reactions. This is a social good that we do not, generally, consider to be research, although the activities are conducted for the purpose of creating generalizable knowledge upon which to base public health decisions. 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, interactions, compliance, or efficacy. These investigations may be identified as research, and they may be required to undergo ethics review board review. However, the difference between these two types of activities can be difficult to demarcate.

HUMAN SUBJECTS

The Common Rule, a set of US Federal regulations that govern research ethics, defines a "research subject" as "a living individual, about whom an investigator conducting research obtains either: 1) data through intervention or interaction with the individual, or 2) identifiable private information." For pharmacoepidemiologists, the key issue 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 key 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 and at institutions that 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. 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. Therefore, the Common Rule serves as *de facto* law governing research at most research institutions in the US and offers a reasonable working definition.

ETHICS REVIEW BOARDS

The consensus that scientists, and science, could benefit from independent review of research protocols is an idea that first appeared in the World Medical Association's Declaration of Helsinki in 1964. The Declaration recommends that the committee be responsible for "consideration, comment and guidance," but does not define the committee's authority to approve or reject protocols that it finds unacceptable. These recommendations have been taken up rapidly, and review boards have become widespread. Their authority has been clarified as well, and the committees typically have the power to review and reject all research that takes place in their institution, region, or nation or in which their investigators are involved.

Over the past 30 years, ethics review boards have become central to the practice of research. The review boards are formed at either the institution, regional, or national level and are often appointed by professional organizations or government agencies. The board is typically composed of scientific expertise in addition to expertise representing law, ethics, or another non-science discipline and at least one member who is not affiliated with the institution. The purpose of these requirements is to introduce accountability to society and minimize conflicts of interest between scientists who act as research reviewers and researchers.

A committee, referred to as an Institutional Review Board (IRB), is required to review all research that is funded by all Federal government branches that have signed on to the "Common Rule." In most other countries, research regulations are not limited by provisions regarding funding but, instead, apply to all research conducted in that country.

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 research may be exempt. 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 ethics board review is any project that does not involve human subjects. 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 (see below for definition) and the research involves "existing data," which means a retrospective analysis of records that exist as of the date the research is proposed. Most European nations have similar provisions for expediting the review of research that poses no more than minimal risks to subjects. Internationally, there is some disagreement about whether pharmacoepidemiologic research should require review. The Royal College of Physicians would not require review, while the Council for International Organizations of Medical Sciences (CIOMS) recommends ethics board review for all research.

PRIVACY AND CONFIDENTIALITY

Privacy, in the setting of research, refers to each individual's right to be free from unwanted inspection of, access to, or physical manipulation of records and documents containing personal information. In the case of epidemiology research in particular, 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. This is a legal way of considering a right to privacy, but privacy is also a precondition for social interaction and cooperation because it allows and requires a degree of trust.

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. 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 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, 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. Communitarianism (the perspective of a community created by voluntary association) argues that the good of the individual is inextricably tied to the public good. 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. Liberalism, or a rights-based individualism, believes in that 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.

Regulations in the US and elsewhere 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. The following sections will discuss both of these key arguments.

MINIMAL RISK

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. 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." 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."

INFORMED CONSENT

Perhaps the most disturbing feature of many of the research scandals in recent history has been the total disregard for informed consent. Every nation which has addressed the subject recognizes that subjects, or, for incompetent patients, their surrogates, are to be told about the nature of research and alternatives to participation and to have the chance to volunteer to participate. 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.

The US regulations, while not universal, 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 without penalty.

The Common Rule requires that written informed consent be obtained in most research situations, with 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 signed consent is the only link the investigator has between personal data and the subject's identity. Second, informed consent can be entirely waived if the research meets four conditions:

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;
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. This view follows from strict adherence to some research ethics codes, and 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. The first is to assure that the ethics board review 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.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGIC RESEARCH

There are several procedures available that can protect patient confidentiality. Patients can:

1. provide universal consent to allow the medical record to be used for research;
2. provide consent to use specified portions of the medical record for research;
3. be contacted and consented for each research protocol.

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 invalid research is inherently unethical, as no risk to individuals is acceptable if there is no benefit to society.

First, there is reason for skepticism about whether these interventions 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 case-control study of patients with and without breast cancer). Although these concerns cover very different issues, 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, and research should not be done if it cannot answer the hypothesis it claims to test. In pharmacoepidemiology studies that use archival records, methods that allow patients to control who has access to their 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. 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 (see Case Example 19.1).

CASE EXAMPLE 19.1: POTENTIAL EFFECT OF AUTHORIZATION BIAS ON MEDICAL RECORDS RESEARCH

Background

- Changes in Minnesota law mandated prior authorization for use of medical records for medical record-based research. The law stated that two requests for authorization be sent to the patient. If the patient did not respond to the request, an implied authorization became effective 60 days after the second mailing. The problem is that restricted access to medical records may limit the ability to conduct valid observational studies because the need to have patients authorize the use of their medical records may create an authorization bias (those who authorize may be different from those who do not, in ways that relate to exposure and/or study outcome).

Question

- What are the characteristics of persons refusing authorization for the use of medical records for research purposes?

Approach

- Investigators compared demographic, diagnostic, and utilization characteristics of patients who provided authorization to those who refused authorization among patients who had received medical care at Mayo Clinic Rochester.

Results

- Overall refusal (nonauthorization) rate was 3.2%; nonresponse was considered as authorization (under the law). When nonresponses were considered refusals, the overall refusal rate increased to 20.7%.
- Patients <60 years old were more likely to refuse than older patients (5.4% versus 1.2%, $p < 0.001$).
- Patients with prior diseases considered sensitive (mental health, infectious disease, and reproductive disorders) were more likely to refuse authorization.

Strength

- The study assessed actual response to requests to release medical records for research purposes.

Limitations

- The study does not determine why participants refused and does not follow up on nonresponders to the invitation to release medical records for research to determine why they did not provide authorization.
- The Mayo practice is a unique population and may be more likely to see the value of medical records research. Hence, this study may underestimate the impact of authorization bias.

Summary Points

- Legislative changes designed to promote control over medical records can affect the validity of medical record-based studies.
- The systematic difference between those who provide authorization and those who refuse authorization may create an authorization bias.
- Further research is necessary to address the privacy concerns of patients with sensitive diseases to ensure that adequate safeguards are in place to promote participation.

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 only among those with a variable which modifies the association between the study medication and the outcome. If the presence of that variable is shielded, the overall study results may show only a

low-level association or no association at all between the study drug and the outcome, missing a true risk.

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 nonrandom fashion. 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.

As noted elsewhere in this text, the number of studies using archival records will likely continue to 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.

CURRENTLY AVAILABLE SOLUTIONS

These methodological 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 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. The available solutions to the methodological challenges outlined in the previous section 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.

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 in October, 1998 covers all information that is either directly identifiable or information from which identity could be inferred. 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. Notably, however, member states may grant deviations from some provisions of the Directive for activities of substantial public interest. 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 was developed, and went into effect in 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 to be covered entities (e.g., health systems).

Of specific interest for pharmacoepidemiologists are the strategies for protecting confidentiality and enabling researchers to access existing data sets. Under the new 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 would have to be removed. Of course, 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 to be identifiers 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.

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, but the code itself cannot be disclosed. 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 pharmacoepidemiology studies. Additionally, the Privacy Rule preserves 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.

There are also opportunities to improve the ethics board review process. Ethics review varies widely from country to country, and there may even be differences within one country. There is general agreement that protocol review by ethics review boards is valuable. However, there is considerably less agreement about what kinds of pharmacoepidemiologic research require this review and about the review criteria themselves. 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. 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 pharmacoepidemiology investigators, challenges that may be magnified in the case of multicenter research that crosses international borders.

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 pharmacoepidemiology investigators. These should be the focus of future efforts to harmonize research regulations and set minimum standards for ethics review board competency and funding. 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, there is debate over how ethics review boards and investigators should balance ethical and methodological 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 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 epidemiology research are not the usual health risks of research that can be balanced against the potential health benefits of research. 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. 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. The implicit argument for this exception is that these social programs offer clear and evident value, and 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 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 pharmacoepidemiology 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.

Other studies may be done for private companies or organizations following rigorous methodological standards but where the findings would not be made public or shared with anyone outside the sponsoring organization. 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, or even to society at large.

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 performing these in a manner that protects indi-

viduals' right to privacy and confidentiality. 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 to be lost.

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 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.

Key Points

- Much like all human subjects research, the risks involved in pharmacoepidemiology studies span from risks that require an investigator to obtain subjects' informed consent to risks that are accepted as part of public health surveillance and do not require additional oversight or protection.
- Violation of privacy and confidentiality is the chief risk in pharmacoepidemiology studies.
- Ethics boards and investigators must gain the knowledge and expertise to review the risks of pharmacoepidemiology studies in order to develop subject protections that are appropriate to the study's methods.

SUGGESTED FURTHER READINGS

- Andrews E. Data privacy, medical record confidentiality, and research in the interest of the public health. *Pharmacoepidemiol Drug Saf* 1999; **8**: 247–60.
- Beauchamp TL, Cook RR, Fayerweather WE, Raabe GK, Thar WE, Cowles SR, Spivey GH. Ethical guidelines for epidemiologists. *J Clin Epidemiol* 1991; **44**: 151S–69S.
- Brody BA. *The Ethics of Biomedical Research: An International Perspective*. New York: Oxford University Press, 1998.
- Caplan AL. Twenty years after. The legacy of the Tuskegee Syphilis Study. When evil intrudes. *Hastings Cent Rep* 1992; **22**: 29–32.
- Department of Health and Human Services, Protection of Human Subjects. Title 45 Part 46: Revised. Code of Federal Regulation; June 18, 1991.

- Freedman B. Scientific value and validity as ethical requirements for research: a proposed explication. *IRB* 1987; **9**: 7–10.
- Guidelines for good pharmacoepidemiology practices. *Pharmacoepidemiol Drug Saf* 2005; **14**: 589–95.
- 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.
- Last JM. Guidelines on ethics for epidemiologists. *Int J Epidemiol* 1990; **16**: 226–9.
- Mann RD, Bertelsmann A. The issue of data privacy and confidentiality in Europe—1998. *Pharmacoepidemiol Drug Saf* 1999; **8**: 261–4.
- 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.
- Olsen J, Breart G, Feskens E, Grabauskas V, Noah N, Olsen J, Porta M, Saracci R. 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.
- The Nuremberg Code. Reprinted in: Brody, H. *The Ethics of Biomedical Research. An International Perspective*. New York: Oxford University Press, 1998; p. 213.

The Use of Randomized Controlled Trials for Pharmacoepidemiology Studies

The following individuals contributed to editing sections of this chapter:

SAMUEL M. LESKO and ALLEN A. MITCHELL

Slone Epidemiology Center, Boston University, Boston, Massachusetts, USA.

INTRODUCTION

Because they provide unbiased estimates of effect, randomized controlled trials (RCTs) are considered the gold standard for demonstrating the effectiveness of a new medication (see Chapter 2). While RCTs are generally used to evaluate beneficial drug effects (see also Chapter 21), 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 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 relatively small differences in the risk of common adverse events or to estimate reliably the risk of rare events. 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 are not generally conducted

in the postmarketing setting. The authors' search for the best method to assess the risk of serious but rare adverse reactions to pediatric ibuprofen and the resulting experience serves as the basis for this chapter (see Case Example 20.1) and may prompt others to consider randomized trials for the postmarketing assessment of drug safety.

CASE EXAMPLE 20.1: THE RISKS OF SHORT-TERM IBUPROFEN USE IN CHILDREN

Background

- The use of nonsteroidal anti-inflammatory drugs is associated with an increased risk of gastrointestinal bleeding and renal failure in adults.
- In 1989, ibuprofen suspension (an NSAID) was approved for use in children by prescription only.
- The risk of rare but serious adverse events among children treated with ibuprofen suspension must be documented before this medication can be considered

(Continued)

for a switch from prescription to over-the-counter use in children.

- Confounding by indication is likely in observational studies of prescription ibuprofen use in children.

Question

- Is the use of ibuprofen suspension in children associated with an increased risk of rare but serious adverse events?

Approach

- Conduct a large simple randomized trial of ibuprofen use in children.
- A randomized trial involving nearly 84 000 children 12 years of age and younger with a febrile illness was conducted.

Results

- The risk of rare but serious adverse events (hospitalization for gastrointestinal bleeding, acute renal failure, anaphylaxis and Reye syndrome) was not significantly greater among children treated with ibuprofen compared to those treated with acetaminophen.

Strengths

- The large sample size allowed evaluation of rare events.
- Randomization effectively controlled for confounding, including confounding by indication.

Limitations

- The use of an active control treatment (acetaminophen) precludes using these data to compare the risk of ibuprofen to that of placebo in febrile children.
- Because medication exposure was limited to the duration of an acute illness, this study cannot be used to assess the risk of long-term ibuprofen use in children.

Summary Points

- When confounding by indication is likely, a randomized controlled trial may be the only study design that will provide a valid estimate of a medication's effect.
- Large, simple, randomized controlled trials can be successfully conducted to evaluate medication safety.
- By keeping the study simple, it is possible to conduct a large, practice-based study and collect data that reflect current ambulatory medical practice.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Pharmacoepidemiologic methods are used to quantify risks and benefits of medications that could not be adequately evaluated in studies performed during the premarketing phase of drug testing. While this chapter considers only the assessment of the risks of medications, the principles involved also apply to the postmarketing evaluation of the benefits of medications (see also Chapter 21).

As noted in Chapters 1 and 3, 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 1 per 1000 treated patients. Modest differences in risk of non-life-threatening adverse events can be of substantial public health importance, particularly if the medication is likely to be used by large numbers of patients. If there are post-licensing questions about the safety of a drug, large observational studies are typically used to satisfy the sample sizes needed to identify (or rule out) the relevant risks. 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 (see Chapters 2 and 16).

Weak associations deserve particular attention with respect to uncontrolled confounding. 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 infer causality because such an association, 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. Thus, it is advisable to use extreme caution in making causal inferences from small relative risks derived from observational studies. When there is concern about residual confounding prior to embarking on an observational study, one may wish to consider using a non-observational study design.

METHODOLOGIC PROBLEMS TO BE SOLVED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Confounding by indication (also referred to as indication bias, channeling, confounding by severity, or contraindication bias) may be a particular problem for postmarketing drug studies. According to Slone *et al.* (1979), confounding by indication exists when “patients who receive different treatments . . . differ in their risk of adverse outcomes, independent of the treatment received.” 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 (see also Chapters 16 and 21). 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 often is not easily done.

Confounding by indication is a particular concern in a number of settings (see Chapters 16 and 21). In general, observational studies are most informative when patients receiving different medications are similar with respect to their risks of adverse events. When there is a single therapy for an illness, and all patients receive that 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. 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, as can a large relative risk if the outcome is rare.

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 consider performing 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 RCTs

As noted above, 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 equalize the distributions of confounding factors, whether they are known or unknown. Therefore, the assigned treatment is the most likely explanation for any observed difference between treatment groups in the clinical outcomes (improvement in the illness or the occurrence of adverse clinical events). By definition, participants in observational studies are not assigned treatment at random. In clinical practice, 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, any of 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 among 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 (see Chapter 3). 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 common clinical outcomes or large differences in uncommon ones.

Blinding

Blinding is used to minimize detection bias, and is particularly important where the outcome is 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 used. 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 an NSAID rather than acetaminophen. Thus, follow-up data collection will only be unbiased if both parties (patient and investigator) are unaware of the treatment assigned. Blinding may be difficult to achieve and maintain, particularly if either the study or control medication produces specific symptoms (i.e., side effects) or easily observable physiologic effects (e.g., nausea or change in pulse rate).

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. Further, it may be difficult to maintain blinding in placebo-controlled studies, as noted above. Studies employing an active control typically utilize usual 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 a 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 general safety and therefore must collect extensive data on symptoms, physical signs, and laboratory evaluations.

Data Analysis

In observational studies, data analyses may be quite complex because of the need to adjust for potential confounders. In contrast, analysis of the primary hypothesis in many clinical trials is straightforward and involves a comparison 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 complexity.

Generalizability of Results

The usual clinical trial conducted during the premarketing evaluation of a drug almost always involves highly selected patients; as a consequence, the results of the trial may not be generalizable to the large numbers of patients who may use the medication after licensing. Observational studies offer an advantage in that they can reflect the real-world experience of medication use and clinical outcomes, and because their modest costs permit studies involving large numbers of patients.

LIMITATIONS OF RCTS

Methodologic strengths notwithstanding, there are several features of the classic RCT that limit its use as a postmarketing study design. First, it may be unethical to conduct a study in which patients are randomly assigned a potentially harmful treatment. For example, an RCT to test the hypothesis that cigarette smoking increases the risk of heart disease would not be acceptable. Second, 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. 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.

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, and the volume and complexity of data collection can be kept to a minimum. The US Salk vaccine trial of the 1950s is an early example of a very large trial. More recently, large randomized trials have been used to test the efficacy of therapeutic interventions, especially in cardiology, or to evaluate dietary supplements or pharmaceuticals for primary prevention of cardiovascular disease and cancer. 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. 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 both known and unknown factors, and the large study size provides the power needed to evaluate small risks of common events as well as large risks of rare events.

How Simple Is Simple?

Yusuf *et al.* (1984) suggest that very large randomized studies of treatment-related mortality collect only the participants' vital status at the conclusion of the study. Because the question of drug safety frequently concerns outcomes less severe than mortality, these ultra simple trials may not be sufficient. Hasford (1994) has suggested an alternative in which "large trials with lean protocols" include only

relevant baseline, follow-up, and outcome data. 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 a function of the number of events observed during the course of the study, which in turn is determined by the incidence rate for the event, the sample size, and the duration of observation or follow-up (see Chapter 3). 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 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. However, power is not the only factor to consider. For example, while an elderly population may be at high risk for gastrointestinal bleeding or cardiovascular events, a study limited to this group may lack generalizability and would not provide information on the risk of these events in younger adults or children.

Data Elements

The data collection process can be kept simple by examining primary endpoints that are objective, easily identified, and verifiable. Because confounding is controlled by randomization, data on potential confounders need not be collected. Rather, a few basic demographic variables can be collected at enrollment in order to characterize the population and to confirm that randomization was achieved.

Data Collection

The data collection process itself can be simplified; follow-up data can be collected by mailed questionnaires or telephone interviews conducted directly with the study participants. Because the study will involve clear and objective outcomes (see below), which can be confirmed by medical record review or other means, self-report by the 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 LSTs conducted among subscribers of a large health maintenance organization) or vital status records for fatal outcomes (e.g., the US National Death Index).

The primary advantage of simplicity is that it allows very large groups of study participants to be followed at

reasonable cost. However, 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 SIMPLE RANDOMIZED TRIAL APPROPRIATE?

LSTs are appropriate when all of the conditions in Table 20.1 apply.

Important Research Question

Although a simple trial will cost less per subject than a traditional clinical trial, the total cost of a large study (in money and human resources) will still be substantial. The cost will usually be justified only when there is a clear need for a reliable answer to a question concerning the risk of a serious outcome. 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. On the other hand, 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.” Originally described by Gray *et al.* (1995) as a simple criterion to assess subject eligibility in LSTs, 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 would apply this principle to evaluate if 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

Table 20.1. Conditions appropriate for the conduct of a large simply 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.
 - (3b) The relative risk is small, regardless of the absolute risk.
 - (4) Important effect modification (interaction) is unlikely.
-

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 superior therapeutic benefits, but it may be unknown whether the risks of side effects outweigh this 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.

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 from any source). 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. 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 any subgroup. 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 AN LST FEASIBLE?

LSTs are feasible when all of the conditions in Table 20.2 are met.

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?"

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 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 and easy to define ("simple"), identify, and recall. An example might include hospitalization for acute gastrointestinal bleeding. Study participants may not correctly recall the details of a hospital admission, or even the specific reason for 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

Table 20.2. 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.
-

interviews, or linkage with public vital status records). On the other hand, clinical outcomes that can be reliably detected only by detailed in-person interviews, physical examinations, or extensive physiologic testing may not be amenable for study in simple trials.

Cooperative Population

Particularly in LSTs, a cooperative and motivated study population will greatly increase 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. 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.

LOGISTICS OF CONDUCTING AN LST

An LST may be appropriate and feasible, but it will only succeed if all logistical aspects of the study are kept simple as well. In general, LSTs 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) can participate by identifying or recruiting eligible patients and obtaining informed consent. Because success depends on the cooperation of multiple health care providers and a large number of patients, it may be best to limit the demands placed on each practitioner (or his/her clinical practice).

To facilitate patient recruitment and to maximize generalizability of the results, minimal restrictions should be placed on patient eligibility. 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 in triplicate with one copy kept on file by the enrolling collaborator, one given to the study participant, and one forwarded to the data coordinating center by mail or facsimile. Registration of study subjects can also be accomplished online using a secure Internet (or dialup) connection to the coordinating center, which allows for immediate confirmation of eligibility and randomization.

Substantial bias can be introduced if either physician or patient can choose not to participate after learning (or guessing) which treatment the patient has been assigned. 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 compliance with the treatment regimen. To minimize dropouts, 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. During the run-in period, eligible subjects are given a "test" medication to assess compliance with the protocol; patients who cannot comply are withdrawn from the trial. Those who remain are likely to be highly compliant, 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. As a consequence, however, the data cannot be used to assess the risk of gastrointestinal bleeding following aspirin use. In addition, excluding patients who do not adhere to therapy or have side effects during the run-in period can limit the generalizability of the study.

Importance of Complete Follow-Up

Because dropouts and losses to follow-up may not be random but may be related to adverse treatment effects, it is important to make every effort to obtain follow-up data on all subjects. A study with 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 it is too short, important outcomes may be missed (i.e., some conditions 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 an observational cohort study because of selective dropouts in either or both of the treatment arms. Beyond choosing a motivated and interested study population, the investigators can minimize losses to follow-up by maintaining regular contact with all study participants. Regular mailings of supplies of medication, a study newsletter, or email reminders can be helpful, and memory aids such as

medication calendar packs or other devices may help maintain compliance with chronic treatment schedules.

Follow-Up Data Collection

Follow-up data collection is the responsibility of the central study staff. Busy health care providers cannot be expected to commit the time required to obtain systematically even minimal follow-up data from large numbers of subjects. However, the clinician who originally enrolled the subject may be able to provide limited follow-up data (e.g., vital status) or a current address or telephone number for the occasional patient who would otherwise be lost to follow-up. A questionnaire delivered by mail, supplemented by telephone interviews when needed, is effective. The response rate will likely be greatest if the questions are both simple and direct and minimal time is required to complete the questionnaire. Medical records can be reviewed to verify important outcomes, such as rare adverse events, and 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 blood or other 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 simple 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 test the randomization procedure; any material differences between treatment groups suggest an imbalance despite randomization. As noted above, it is 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 among the different groups. 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 will 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 potentially at risk for adverse outcomes, it is appropriate to monitor the accumulating data during the study. A 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, should conduct periodic reviews of the data using an appropriate analysis procedure.

THE FUTURE

With accelerated approval of new medications and rapid increases in their use, we may see a greater need for large, randomized postmarketing studies to assess small differences in risk. This is particularly the case for drugs considered for over-the-counter 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 by much larger and 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. Improvements in the efficiency with which such trials can be carried out may lead to their increased use.

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 (see Chapters 11 and 12). 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 would have to be provided that events were not missed by patients presenting to out-of-plan providers. In theory, it may be possible to conduct such a “hybrid trial,” but to our knowledge such a trial has not been attempted.

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 new 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, although it would at least be possible to accurately estimate the absolute risk of important events among exposed subjects. Where feasible, patients could be randomized to receive different doses, and a dose–response relationship could be sought.

It is clear that very large simple controlled trials of drug safety can be successfully carried out. It is less clear, however, how frequently the factors that indicate the need for a very large trial (Table 20.1) will converge with those that permit such a trial to be carried out (Table 20.2). As pharmacoepidemiologists become more familiar with LSTs, we may see more of them being conducted, and new methods of subject recruitment and more efficient sources of follow-up data are likely to be developed.

Key Points

- Randomization usually controls for confounding, including confounding by indication.
- A large study allows assessment of small to modest associations of common events and large associations with rare events and assures that randomization produces balanced treatment groups.
- Large randomized controlled trials are feasible if data collection is kept simple and outcome events are objective and verifiable.

SUGGESTED FURTHER READINGS

- Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancer in male smokers. *N Engl J Med* 1994; **330**: 1029–35.
- DeMets DL. Data and safety monitoring boards. In: Armitage P, Colton T, eds, *Encyclopedia of Biostatistics*. Chichester: John Wiley & Sons, 1998; pp. 1067–71.
- Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah 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.
- Francis T Jr, Korn 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.
- Frommer DJ. Case–control studies of screening. *J Clin Epidemiol* 1988; **41**: 101.
- Gray R, Clarke M, Collins R, Peto R. Making randomized trials larger: a simple solution? *Eur J Surg Oncol* 1995; **2**: 137–9.
- Hasford J. Drug risk assessment: a case for large trials with lean protocols. *Pharmacoepidemiol Drug Saf* 1994; **3**: 321–7.
- 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.
- 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.
- 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.
- ISIS-1 (First International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous atenolol among 16,027 cases of suspected acute myocardial infarction: ISIS-1. *Lancet* 1986; **ii**: 57–66.
- Knox G. Case–control studies of screening procedures. *Public Health* 1991; **105**: 55–61.
- 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.
- Lesko SM, Mitchell AA. An assessment of the safety of pediatric ibuprofen: a practitioner-based randomized clinical trial. *JAMA* 1995; **273**: 929–33.
- 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.
- Morrison AS. Case definition in case–control studies of the efficacy of screening. *Am J Epidemiol* 1982; **115**: 6–8.
- O’Brien PC. Data and safety monitoring. In: Armitage P, Colton T, eds, *Encyclopedia of Biostatistics*. Chichester: John Wiley & Sons, 1998; pp. 1058–66.
- Paul R, Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.

- Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med* 1994; **331**: 394–8.
- 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.
- Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. A policy perspective. *Ann Intern Med* 1979; **90**: 257–61.
- Strom BL, Miettinen OS, Melmon KL. Postmarketing studies of drug efficacy: when must they be randomized? *Clin Pharmacol Ther* 1983; **34**: 1–7.
- Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? *Am J Med* 1984; **77**: 703–8.
- Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? *Am J Med* 1985; **78**: 475–80.
- Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; **3**: 409–20.

The Use of Pharmacoepidemiology to Study Beneficial Drug Effects

Edited by:

BRIAN L. STROM

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 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 complete information about the most clinically important drug toxicity. 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.

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, nor was the skeletal muscle pain associated with use of HMG-CoA reductase inhibitors known. 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. 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 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. 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 nonsteroidal anti-inflammatory drugs have the same beneficial effects; some data suggest such an effect while other data suggest an increased risk of cardiovascular events. Secondly, 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 nonexperimental 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 22.

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 academic researchers (see Chapter 15 for a discussion of the validity issues involved in measuring such outcomes); they can be measures of quality-of-life (see Chapter 23), often referred to in the pharmaceutical

industry as "outcomes research"; they can be measures of utility, i.e., global measures of the desirability of certain clinical outcomes (see Chapters 22 and 23); they can be economic outcomes (see Chapter 22); etc. Regardless, the same methodologic issues apply to each.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGIC 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 21.1). Premarketing randomized clinical trials generally provide information on whether a drug can produce at least one beneficial effect. Specifically, premarketing studies generally investigate the efficacy of 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 atypically high, and the patients included are similar to each other in age and sex, usually do not have other diseases, and are taking few, if any, 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

Table 21.1. 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, compliance, other illnesses, drugs taken for this or other illness (including tobacco and alcohol), etc.

Source: Modified from Strom *et al.* (1985).

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 limited data, if any, 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 Food and Drug Administration (FDA) at the time of its regulatory decisions about the marketing and labeling of the drugs involved in these uses. 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 for the specific use at the time of the comparison; 8 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, 8 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.

Drugs other than those in the list of 100 common drug uses, i.e., drug-indication pairs, were sometimes prescribed as treatment for each of the 52 indications included in those 100 pairs. 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 21.2. 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 that were for patients less than age 20 ranged from 0.0% for 43 of the uses to 97%. Yet, many of these uses had not been tested in children prior to marketing. Analogously, only three of the drugs were approved for use in pregnant patients, yet we know that drug use in pregnancy was common, even then.

Thus, this study revealed considerable gaps in the information about beneficial drug effects at the time of drug marketing. These deficiencies in the available information should not be surprising, nor should they be considered inadequacies that ought to prevent the release of the drug to the marketplace. The data needed for clinical decisions are frequently and understandably different from those needed for regulatory decisions. Studies performed prior to marketing perforce are focused predominantly on meeting appropriate regulatory requirements, and only secondarily on providing a basis for optimal therapeutic decisions. 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 2 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

Table 21.2. 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
<i>Regimen</i>			
Dose per unit time	Ibuprofen	Rheumatoid arthritis, osteoarthritis	Daily dosage initially approved proved to be suboptimal
Distribution of dose over time	Furosemide	Congestive heart failure	Efficacy improved by more frequent, smaller doses
Duration	Clonidine Hypoglycemics (e.g., acetohexamide and tolazamide)	Hypertension Diabetes mellitus	Tolerance develops in the absence of a diuretic Tolerance develops in many patients
<i>Indication</i>			
Severity	Metaproterenol	Asthma	Patients with severe illness do not have a response without additional, supplementary therapy
Subcategories	Desipramine	Depression	May vary with endogenous versus exogenous depression
Changes over time	Ampicillin	Otitis media	No longer the drug of choice in some geographic areas due to bacterial resistance
<i>Patient</i>			
Age	Diazepam	Anxiety	A given regimen is more effective in the aged than in the young 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 paradoxical reactions
Other illness	Gentamicin	Infection	Lower doses required in renal failure
<i>Other</i>			
Drugs	Lithium Acetohexamide	Manic-depressive illness Diabetes mellitus	Clearance impaired by diuretics, e.g., furosemide Many drugs interfere, by causing hyperglycemia (e.g., diuretics), displacing drug from binding sites (e.g., nonsteroidal anti-inflammatory drugs), etc.
Diet	Diuretics (e.g., metolazone, furosemide) Lithium	Hypertension Manic-depressive illness	A decrease in sodium intake can improve efficacy Significant sodium depletion or excess can modify renal excretion

Source: Strom *et al.* (1985).

of the other two and, thereby, can create an apparent association or mask a real one. This is discussed in more depth in Chapter 16. Studies of intended drug effects present a special methodologic problem of confounding by the indication for therapy. 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.

Although confounding by the indication is a less common problem for studies of side effects, this is not the case for studies of anticipated beneficial effects. In these studies one would expect the indication to be more closely related to the outcome variable. In fact, the problem presented by confounding by the indication has been thought by some to invalidate nonexperimental approaches to studies of the beneficial effects of drugs. Some have felt that questions of beneficial drug effects can be addressed only by using randomized clinical trials. 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 21.3). First, some questions do not require any comparative (analytic) research for their answer. For these, simple clinical observations, as reported in a case report or case series, can be sufficient. For example, the efficacy and effectiveness of naloxone, used as a narcotic antagonist, is demonstrable simply through the observation of a single patient. Consider a patient comatose from an overdose of methadone. An injection of naloxone results in his prompt awakening. However, 30 minutes later, as the effects of the narcotic antagonist wear off, the patient returns to coma. Another injection of the naloxone results in awakening once more, and then later the coma returns again. This sequence of events represents a convincing demonstration of the drug's ability to have its desired effect. No elaborate studies are needed to make this point. The same would be 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 usually will tend to improve spontaneously. 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.

Recently, researchers have begun to use *propensity scores* towards this end. This is an approach that uses mathematical modeling to predict *exposure*, rather than the traditional approach of predicting *outcome*. This is, essentially,

Table 21.3. 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 (b) Drug effect obvious in a series of patients	Naloxone used for methadone overdose 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 (ii) Gradations in the indication are unrelated to the choice of treatment, or (iii) Gradations in the indication are unrelated to expected outcome, or (iv) Special clinical settings (b) The indication is sufficiently characterizable (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	Anti-Rh (D) immune globulin given to Rh (D) negative mothers who deliver Rh (D) positive newborns to prevent future erythroblastosis fetalis Penicillin used for endocarditis prophylaxis in patients with congenital aortic stenosis who are undergoing tooth extraction Penicillin used to prevent tertiary syphilis, given to patients with an asymptomatic positive serologic test for syphilis Anticoagulants used after myocardial infarctions to prevent death Isoniazid used for tuberculosis prophylaxis in a patient with an asymptomatic positive purified protein derivative
(4) Confounding by the indication exists and is not controllable	Ampicillin used to treat urinary tract infection

Source: Strom *et al.* (1983).

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, based on very recent data, propensity score only has 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.

When questions of intended drug effects do not fall into any of the preceding categories, *confounding by the indication cannot be controlled*. Nonexperimental study designs cannot then be used, or they can only be used to demonstrate qualitatively some degree of beneficial effect. Specifically, if confounding by the indication is such that treated patients would have a worse clinical outcome than untreated patients, yet the outcome observed in treated patients is better than that observed in untreated patients, some degree of confidence that the drug has a beneficial effect can be built. As

an example, patients treated with corticosteroids for status asthmaticus would be expected to be sicker than those not so treated. If patients receiving corticosteroids stop wheezing sooner than those not receiving corticosteroids, corticosteroids would indeed seem to have a beneficial effect. However, if the patients receiving corticosteroids do not stop wheezing sooner than those not receiving corticosteroids, the results of the study are uninterpretable. It is possible that the corticosteroids in fact have no beneficial effect. 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 21.3 and discussed further in Strom *et al.* (1983).

APPLICABILITY OF THE PROPOSED APPROACHES

How commonly are the nonexperimental approaches we have described applicable for the study of beneficial drug effects? A list of the 100 most recently approved new molecular entities as of December 1978 was studied to determine what types of nonexperimental study designs, if any, could be used to evaluate drug effectiveness. After excluding from this list 7 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. Of these drug uses, 89 (67.9%) 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, e.g., 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 7 of the 42 (5.3% of the total), this confounding was not an obstacle to valid nonexperimental research. Most often the validity of the approach rested on the observation that any given physician usually used the drug to treat either all or none of his patients with the indication.

In the remaining 35 of the 42 uses (26.7% of the total), confounding by the indication was judged to be uncontrollable using currently available nonexperimental techniques.

To place these findings in perspective, of the 42 drug uses that required comparative research to evaluate their effectiveness, 30 could not ethically be addressed using a randomized clinical trial and a placebo control. Most of these 30 involved the use of drugs to treat infections or malignancies. In these situations, patients could not ethically be left "untreated," that is assigned to the placebo group.

Studies of the effects of one drug relative to another active drug, of course, gave different results. Formal comparative research was necessary for all 131 drug uses. Nonexperimental studies theoretically could be conducted validly

for 94 of the 131 drug uses (71.8%). Experimental studies would be ethical for all of them.

Of course, judging theoretically that a question of effectiveness is "studiable" by a given technique is not the same as proving that a valid outcome would emerge from such a study. There are many particular details in the actual conduct of such studies that must be addressed on a case-by-case basis. It is, therefore, instructive to examine some specific examples of nonexperimental research into beneficial drug effects.

SPECIFIC EXAMPLES

Estrogens for Prevention of Osteoporotic Fractures

One of the first series of studies of drug effectiveness using rigorous nonexperimental study designs examined whether exogenous estrogens could prevent fractures in postmenopausal women with osteoporosis. 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. Studies of bone density documented that exogenous estrogens prevented the loss of bone density that was associated with the menopause, for as long as the estrogens were continued. 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.

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 nonexperimental studies were performed. Both case-control and cohort designs were used. (See Case Example 21.1). 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. 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 and endometrial cancer, it is reasonable to assume that some physicians might preferentially routinely use the drugs and others might routinely

avoid them. In such a setting, nonexperimental techniques could yield valid results, unaffected by confounding by the indication (category (3)(a)(iv) in Table 21.3). 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 21.3. 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.

**CASE EXAMPLE 21.1:
A NON-EXPERIMENTAL STUDY TO
DETERMINE THE EFFECTS OF ESTROGEN
ON FRACTURES**

Background

- Postmenopausal estrogens have long been recognized as slowing the development of osteoporosis, but had not been shown to be effective in preventing osteoporotic fractures, at least in part because of their infrequency.

Question

- Can postmenopausal use of estrogens prevent hip fractures?

Approach

- Perform a case-control study comparing patients hospitalized with hip fractures to randomly selected population controls, examining the relative prevalence of previous exposure to postmenopausal estrogens.

Results

- The odds ratio associated with use of estrogens was indeed protective.

Strengths

- This infrequent event could be studied with a relatively small, efficient study.

- Population-based choice of controls.

Limitation

- Patients receiving postmenopausal estrogens are different from those not receiving them, as has been confirmed in many studies, leaving open the possibility of residual confounding by indication.

Summary Points

- Case-control studies can be useful to study the effectiveness of treatments in preventing uncommon outcomes.
- Confounding by indication will always remain a residual concern in such studies.

In fact, the three studies that closely examined the comparability of the study groups were able to document that they were not comparable. 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. 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.

A third study used a case-control design to investigate patients admitted to surgical services. 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 nonexperimental studies published since then showed similar results. Since then, the finding that estrogens have a beneficial effect on hip fractures has been confirmed in a massive clinical trial, the Women's Health Initiative.

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. 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 22.

Vaccines

In the past several years, nonexperimental study designs have been widely used to evaluate the efficacy of vaccines. Specifically, case-control studies have been used to explore the efficacy of pneumococcal vaccine, rubella vaccine, measles vaccine, *Hemophilus influenzae* type b polysaccharide vaccine, oral poliovirus vaccine, meningococcus vaccine, Japanese encephalitis vaccine, and BCG vaccine in protecting against tuberculosis, diphtheria toxoid vaccine, and leprosy. Cohort studies have been used to explore the efficacy of *Hemophilus influenzae* type b polysaccharide vaccine, measles vaccine, and pertussis vaccine. 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 the 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, while a study of its safety, conducted by the same authors, used a randomized clinical trial design. In considering the applicability of nonexperimental study designs, the relatively indiscriminate use of such vaccines places the study in category (2) of Table 21.3. 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. Nonexperimental studies of such questions should produce valid results, therefore. Indeed, as is evident from the large number of examples, this is becoming a standard and accepted approach. We refer the interested reader to two methodologic papers on the subtleties of designing nonexperimental studies of vaccine efficacy (Smith *et al.*, 1984; Smith, 1987).

Cancer Screening

Another more recent and frequent use of nonexperimental study designs is to evaluate the efficacy of cancer screening programs. Although this does not directly relate to drugs, the methodological implications are the same, and have been better enunciated than in the pharmacoepidemiology literature. The use of nonexperimental 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 nonexperimental designs. Screening procedures that have been evaluated repeatedly in this fashion include the value of "Pap" smears for cervical cancer and mammography and self-examination for breast cancer. Other studies investigated screening measures for lung cancer and gastric cancer. All of these were case-control studies. Many more have been published since. 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. 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.

Other Examples

Other analogous work using case-control study designs has explored the effectiveness of bicycle safety helmets in preventing face injuries, antibiotic prophylaxis in preventing post-dental infective endocarditis, beta-blockers in preventing mortality in patients with acute myocardial infarction, beta-blockers and incident coronary artery events, etc. Other examples of nonexperimental studies include comparisons of one drug against another with a similar indication, including comparing a brand name drug to a generic drug (see Case Example 21.2).

CASE EXAMPLE 21.2: **A NON-EXPERIMENTAL STUDY** **COMPARING A BRAND NAME DRUG** **TO A GENERIC DRUG**

Background

- Generic drugs are approved for marketing on the basis of studies of bioequivalence, i.e., equivalent bioavailability. However, it is not clear whether bioequivalence assures clinical equivalence, that is, equivalent efficacy and toxicity.

Question

- Is generic drug X clinically equivalent to brand name innovator drug A?

Approach

- Perform a large-scale postmarketing cohort study comparing drug X to drug A.

Results

- The effects of drug X and drug A were indistinguishable, with very tight confidence intervals.

Strengths

- Large sample size.
- Cost-effective study.

- Generically equivalent drugs should be used in comparable populations, limiting the risk of confounding by indication.

Limitation

- Patients treated with a generic drug may not be the same as those treated with the brand name innovator, resulting in residual confounding by indication.

Summary Points

- Studies of relative efficacy are comparing two different active drugs, thereby seeking small differences and requiring large sample sizes.
- Studies of relative efficacy are comparing two different drugs for the same indication, minimizing the risk of confounding by indication.
- Even when studying two different products for the same drug, there may be systematic reasons why the patients receiving the drugs are different, resulting in confounding by indication.

THE FUTURE

Clinicians have long recognized the value of clinical observations and nonexperimental research. Much of our current knowledge about the usefulness of medical interventions is based on information that is nonexperimental. 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 nonexperimental studies, some scientists have insisted that “the randomized clinical trial (RCT) is the only scientifically reliable method for assessment of the efficacy (and risks) of most clinical treatments.” Sackett *et al.* (1985) argue: “to keep up with the clinical literature . . . discard at once all articles on therapy that are not randomized trials.” 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 (1966), 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.” Many investigators are now applying nonexperimental designs to studies of beneficial drug effects. However, careful attention needs to be paid to the possibility of confounding by the indication. Some approaches to this problem are now available, and hopefully more will be available in the future. However, when confounding by indication can be addressed, clinical observations and nonexperimental research can be used. The results of nonexperimental research are unlikely to be as powerful or as convincing as those of experimental research. We are not suggesting that nonexperimental studies be used as replacements for experimental studies. However, when an experimental study is deemed to be unnecessary, unethical, infeasible, or too costly relative to the expected benefits, there frequently is a good alternative.

Key Points

- Anticipated beneficial effects are the desirable effects that are known to be caused by the drug. They represent the reason for prescribing the drug.
- There are considerable gaps in the information about beneficial drug effects available at the time of drug marketing.
- Studies of anticipated beneficial effects suffer from a unique methodologic concern: confounding by the indication for drug therapy.
- Because of confounding by indication, most studies of drug effectiveness should be randomized clinical trials.
- There are special situations when nonexperimental study designs can be used to study drug effectiveness, but they must be selected carefully.

SUGGESTED FURTHER READINGS

- 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.
- Cole P, Morrison AS. Basic issues in population screening for cancer. *J Natl Cancer Inst* 1980; **64**: 1263–72.
- 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.
- 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.
- 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.
- Frommer DJ. Case-control studies of screening. *J Clin Epidemiol* 1988; **41**: 101.
- Knox G. Case-control studies of screening procedures. *Public Health* 1991; **105**: 55–61.
- Hill AB. Reflections on the controlled trial. *Ann Rheum Dis* 1966; **25**: 107–13.
- 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.
- Morrison AS. Case definition in case-control studies of the efficacy of screening. *Am J Epidemiol* 1982; **115**: 6–8.
- Paul R, Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.
- Sackett DL, Haynes RB, Tugwell PT. *Clinical Epidemiology. A Basic Science for Clinical Medicine*. Boston, MA: Little, Brown and Co., 1985.
- Sasco AJ, Day NE, Walter SD. Case-control studies for the evaluation of screening. *J Chronic Dis* 1986; **39**: 399–405.
- Sasco AJ. Lead time and length bias in case-control studies for the evaluation of screening. *J Clin Epidemiol* 1988; **41**: 103–4.
- 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.
- Smith PG. Evaluating interventions against tropical diseases. *Int J Epidemiol* 1987; **16**: 159–66.
- Strom BL, Miettinen OS, Melmon KL. Postmarketing studies of drug efficacy: when must they be randomized? *Clin Pharmacol Ther* 1983; **34**: 1–7.
- Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? *Am J Med* 1984; **77**: 703–8.
- Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? *Am J Med* 1985; **78**: 475–80.
- Weiss NS. Control definition in case-control studies of the efficacy of screening and diagnostic testing. *Am J Epidemiol* 1983; **118**: 457–60.
- 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.
- 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.

Pharmacoeconomics: Economic Evaluation of Pharmaceuticals

The following individuals contributed to editing sections of this chapter:

KEVIN A. SCHULMAN,¹ HENRY A. GLICK,² and DANIEL POLSKY²

¹Center for Clinical and Genetic Economics, Duke Clinical Research Institute, Duke University Medical Center, Durham, North Carolina, USA; ²Leonard Davis Institute of Health Economics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 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. More recently, health care researchers from a variety of disciplines have developed techniques for the evaluation of the economic effects of clinical care and new medical technologies.

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. In this chapter, we briefly review the research methods of pharmacoeconomics and discuss some methodologic issues that have confronted researchers investigating the economics of pharmaceuticals.

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.

Concerns about the cost of medical care in general and pharmaceuticals specifically are being felt in nearly all developed nations. Several 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.

Clinical economics research is being used increasingly by managed care organizations in the United States to inform funding decisions for new therapies. 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 two-fold: (i) understanding the potential effect of a therapy (e.g., whether a new anti-sepsis agent is a new type of antibiotic compound, where a short-term evaluation, efficacy at 14 days, is the appropriate clinical endpoint for analysis, or a life supporting therapy, where a longer-term evaluation, efficacy at 6 or 12 months, is the appropriate clinical endpoint for efficacy assessment), and (ii) understanding the transition from efficacy to efficiency in clinical practice. These challenges span the clinical development spectrum. As we learn more about the potential effect 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.

No single study can possibly provide all interested audiences with complete economic information on 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

New pharmaceuticals are developed in a series of well-defined stages. 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 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 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.

Developing economic data as part of a clinical trial requires integrating pharmacoeconomics into the clinical development process. Economic analysis requires the establishment of a set of economic endpoints for study, 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. Two examples of cost analyses within clinical trials are provided in Case Examples 22.1 and 22.2.

CASE EXAMPLE 22.1: **ECONOMIC EVALUATION OF HIGH-DOSE** **CHEMOTHERAPY PLUS AUTOLOGOUS** **STEM CELL TRANSPLANTATION FOR** **METASTATIC BREAST CANCER**

Background

- A clinical trial of high-dose chemotherapy plus autologous hematopoietic stem cell transplantation

versus conventional-dose chemotherapy in women with metastatic breast cancer found no significant differences in survival between the two treatment groups. Thus, the economic evaluation would provide decision makers with important additional information about the two therapies.

Question

- What were the differences between the two treatment groups with regard to course of treatment and resources consumed?

Approach

- The researchers abstracted the clinical trial records and oncology department flow sheets retrospectively to document resource use.
- Each patient's course of treatment and resource use was analyzed in four phases. Based on these clinical phases, patients were grouped into one of three clinical "trajectories."
- Costs were estimated using the Medicare Fee Schedule for inpatient costs and average wholesale prices for medications.
- Sensitivity analyses examined changes in the discount rate, hospital costs, and the number of cycles of paclitaxel and docetaxel.

Results

- Patients undergoing transplantation used more resources, mostly due to inpatient care.
- The investigators also found differences by clinical trajectory, and these differences were not consistent between treatment groups.
- High-dose chemotherapy plus stem-cell transplantation was associated with greater morbidity and economic costs, with no improvement in survival.
- Results of the sensitivity analyses suggested that the findings were robust even when important cost assumptions varied.

Strengths

- By studying resource use and estimating costs, the authors were able to quantify the economic burden associated with the two treatments.
- The study allowed the investigators to provide novel information about the clinical trajectories of patients with metastatic breast cancer
- The economic evaluation did not place any additional data collection burden on investigators, but yielded important secondary findings.

- Economic evaluation allowed the researchers to quantify the economic burden associated with interventions and to provide additional information on patients' "clinical trajectories."

Limitations

- Collection of resource use data from the clinical trial records may have resulted in underestimation of treatment costs.
- Resource costs were estimated rather than directly observed.

Summary Points

- 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.
- Sensitivity analyses are crucial in studies that rely on numerous estimates and assumptions.
- Economic analysis can provide important additional information to decision makers in cases where no differences were observed between treatment groups on the primary clinical endpoint.

CASE EXAMPLE 22.2: **MULTINATIONAL ECONOMIC EVALUATION** **OF VALSARTAN IN PATIENTS WITH CHRONIC** **HEART FAILURE**

Background

- Clinical investigators found no differences in mortality in a clinical trial of ACE inhibitor with the addition of either valsartan or placebo for patients with heart failure.

Question and Issue

- Economic data were collected prospectively as part of the multinational clinical trial.
- What challenges are presented in a multinational economic evaluation?

Approach

- Resource use data were collected at regular intervals in the case report form.
- Hospital and outpatient cost estimates were based on Medicare reimbursement rates for US patients and on

(Continued)

the estimates of local health economists for patients outside the US (based on national fee schedules or hospital accounting systems).

- Cost estimates were converted to 1999 US dollars using purchasing power parties from the Organization for Economic Cooperation and Development.

Results

- Mean cost of a hospitalization for heart failure was \$423 less for patients in the valsartan group, compared to patients in the placebo group.
- Much of the savings was offset by higher costs for non-heart-failure hospitalizations among patients in the valsartan group, yielding a nonsignificant decrease in inpatient costs of \$193 for patients in the valsartan group.
- Overall within-trial costs, including outpatient costs and the cost of valsartan, were \$545 higher for patients in the valsartan group. Thus, there was no finding of dominance.
- However, the subgroup of patients who received valsartan and were not taking an ACE inhibitor at baselines had \$929 lower costs compared to their counterparts receiving placebo, even after including the cost of valsartan. Thus, in this subgroup, valsartan was the dominant strategy.

Strengths

- Data were collected prospectively.
- The study capitalizes on the randomized design of the clinical trial to detect unbiased treatment effects.
- The study demonstrates the relative shift in costs incurred caused by the treatment effect of valsartan.

Limitation

- Practice patterns, resource use, and costs can vary significantly between countries, affecting the generalizability of the findings.

Summary Points

- In multinational economic evaluations, methods of cost-estimation are available that maintain the relation between relative costs and resource use within countries.
- Additional evaluations may be necessary to account for country-specific practice patterns, cost differences, and sociopolitical context.
- Treatment strategies can still be considered dominant even if they increase overall costs.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGIC RESEARCH

TECHNIQUES OF CLINICAL ECONOMICS

In considering economic analysis of medical care, there are three dimensions of such analysis with which readers should become familiar—type of analysis, perspective of the analysis, and types of costs (Figure 22.1).

TYPES OF ANALYSIS

There are three general types of economic analysis of medical care—cost–benefit, cost–effectiveness, and cost identification. Cost–benefit analysis 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. Cost–effectiveness analysis provides an approach to the dilemma of assessing the monetary value of health outcomes as part of the evaluation. Cost identification analysis is a less complex approach than cost–benefit or cost–effectiveness analysis and simply enumerates the costs involved in medical care; it ignores the outcomes that result from that care. This approach is appropriate only if treatment outcomes or benefits are equivalent for the therapies being evaluated. We now provide details for the most common type of economic analysis in medical care, cost–effectiveness analysis

Cost-Effectiveness Analysis

In cost–effectiveness analysis, cost generally is calculated in terms of dollars spent, and effectiveness is determined independently and may be measured only in clinical terms using any meaningful clinical unit, such as number of lives saved, complications prevented, or diseases cured. Health outcomes can also 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 (e.g., dollars per year of life saved). When several outcomes result from a medical intervention, cost–effectiveness analysis may consider these outcomes together only if a common measure of outcome can be developed.

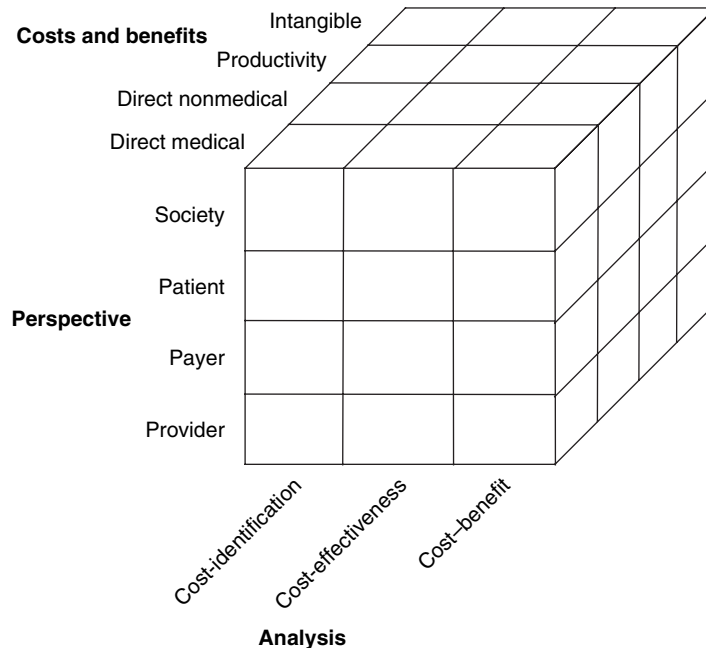


Figure 22.1. The three dimensions of economic evaluation of clinical care (from Bombardier and Eisenberg, 1985).

Cost-effectiveness analysis compares a treatment's incremental costs and incremental effectiveness, resulting in an estimate of the additional effect per additional treatment dollar spent. Programs that cost less and demonstrate improved or equivalent treatment outcomes are said to be dominant and should be adopted. Programs that cost more and are more effective should be adopted if both their cost-effectiveness and incremental cost-effectiveness ratios fall within an acceptable range. If there is a budgetary constraint, this must also be factored into the adoption decision. Programs that cost less and have reduced clinical outcomes may be adopted depending upon the magnitude of the changes in cost and outcome. Programs that cost more and have worse clinical outcomes are said to be dominated and should not be adopted.

Sensitivity Analysis

Most cost-effectiveness studies require large amounts of data that may vary in reliability, validity, or the effect on the overall results of the study. Sensitivity analysis is a set of procedures in which the results of a study are recalculated using alternate values for some variables to test the sensitivity of the conclusions to these altered specifications. 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

Economists consider three types of costs: direct, productivity, and intangible. Direct medical costs usually are associated with monetary transactions and are incurred during the provision of care, such as payments for purchasing a pharmaceutical product, payments for physicians' fees, or purchases of diagnostic tests. Because charges may not accurately reflect the resources consumed, accounting or statistical techniques may be needed to determine direct costs.

Direct nonmedical costs are incurred because of illness or the need to seek medical care. They include, for example, the cost of transportation to the hospital or physician's office, the cost of hotel stays for receiving medical treatment at a distant medical facility, and the cost of special housing. Direct nonmedical costs, which are generally paid out of pocket by patients and their families, are just as much direct costs as are expenses that are typically covered by third-party insurance plans.

If direct costs increase with increasing volume of activity, they are described as variable costs. However, if the same

costs are incurred regardless of the volume of activity, they are described as fixed costs. For example, the paper used in an electrocardiogram machine is a variable cost, since a strip of paper is used for every tracing. However, the machine itself is a fixed cost since it must be purchased whether one tracing is needed or many are performed. Of course, fixed costs are fixed only within certain bounds. A very large increase in activity will require the purchase of another piece of equipment. Even the fixed cost of a hospital's building is only fixed within certain limits of activity and a certain time frame. Still, for the purposes of most decisions in clinical practice, costs can be considered fixed or variable.

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.

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 and are often omitted in clinical economics research.

PERSPECTIVE OF ANALYSIS

The third dimension of economic analysis is the perspective of the analysis. Costs and benefits can be calculated with respect to society's, the patient's, the payer's, and the provider's points of view. The perspective determines which costs and benefits are included. Depending on perspective, the value placed on resources used may not necessarily reflect society's value. 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 from the perspective of the patient.

Because costs will differ depending on the perspective, the economic impact of an intervention will be different from different perspectives. These differences are important in understanding the various dimensions of an adoption decision, but to make comparisons across 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. If an intervention is not good value for money from the societal perspective, it would not be a worth-

while intervention for society, even if the intervention may have economic advantages for other stakeholders.

METHODOLOGIC ISSUES IN THE PHARMACOECONOMIC ASSESSMENT OF THERAPIES

The basic approach for performing economic assessments of pharmaceuticals has been adapted from the general methodology for cost-effectiveness analysis. These methods have been well developed in medical technology assessment and in other fields of economic research. However, there remain a number of methodological issues that confront investigators in economic evaluations of pharmaceuticals. This section reviews some of these issues in the design, analysis, and interpretation of pharmacoeconomic evaluations.

CLINICAL TRIALS VERSUS COMMON PRACTICE

Clinical trials are useful for determining the efficacy of therapeutic agents. However, their focus on efficacy rather than effectiveness 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 as an endpoint in a clinical trial is the performance of routine testing to determine the presence or absence of a study outcome.

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. 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.

Second, protocol-induced testing may lead to the detection of adverse drug effects that would otherwise have gone undetected. 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 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. This potential bias would tend to lower the overall costs of care observed in the trial compared to usual care.

Fourth, 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 generally will increase the costs of diagnostic testing in the trial. 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. This could result in a bias in the study design if there are differences in the amount of services provided to patients in the treatment and control arms of a trial.

Seventh, patients in trials often are carefully selected and 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 many kinds of patients with specific clinical syndromes. These exclusions further limit the generalizability of the findings.

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. Clinical economics must explicitly recognize the 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.

Possible Solutions

One possible solution to this problem is the inclusion of a "usual care" arm of the clinical trial. In such a 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.

A second method that has been used to overcome these problems 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), possibly leaving a biased sample 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.

ISSUES IN THE DESIGN OF PROSPECTIVE PHARMACOECONOMIC STUDIES

Despite their difficulty, prospective pharmacoeconomic studies are often our only opportunity to collect and analyze information on new therapies before decisions are made concerning insurance reimbursement and formulary inclusion for these agents. We now address some issues that arise in the design of these studies. Table 22.1 outlines the steps required for developing an economic analysis plan.

Sample Size

Often those setting up clinical trials focus on the primary clinical question when developing sample-size estimates. They fail to consider the fact that the sample required to

Table 22.1. 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
-

address the economic questions posed in the trial may differ from that needed for the primary clinical question. 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. Power calculations can be performed, however, to determine the detectable differences between the arms of the study given a fixed patient population and various standard deviations around cost and patient preference data.

Participation of Patients

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 the possibility of patient selection bias in the analysis of clinical endpoints. However, given the low rates of refusal to the release of financial information, a single consent form should be considered for all trial data. The single consent would avoid the possibility of selection bias in the economic endpoints relative to the clinical endpoints.

Data Collection

While some prospective data collection is required for almost all pharmaco-economic studies, the amount of data to be collected for the pharmaco-economic evaluation is the subject of much debate. 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.

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. There is a growing literature that addresses the transferability of a study's pooled results to subgroups.

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.

Measurement and Modeling in Clinical Trials

The types of data available at the end of the trial will depend upon the trial's sample size, duration, and clinical endpoint. There are two categories of clinical endpoints considered in pharmacoeconomic analysis: intermediate endpoints and final endpoints. An intermediate endpoint is a clinical parameter, such as systolic blood pressure, which varies as a result of therapy. A final endpoint is an outcome variable, such as change in survival, or quality-adjusted survival, which is common to several economic trials, and allows for comparisons of economic data across clinical studies.

The use of intermediate endpoints 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 endpoints are most appropriate in clinical research if they have been shown to be related to the clinical outcome of interest, as in the Framingham Heart Study, which used changes in blood cholesterol levels to demonstrate the efficacy of new lipid lowering agents.

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, some 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.

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. A one-time effect model assumes that the clinical benefit observed in the trial is the only clinical benefit received by patients. 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 treatment outcome.

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. The analysis plan should describe the study design 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 endpoints, and describe how the endpoints 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.

Also, the analysis plan should identify the statistical methods that will be used and how hypotheses will be tested. 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 (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. 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). The recent proposal of the generalized linear model promises to improve the predictive power of multivariable analyses.

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.

One of the most dependably accurate methods for deriving 95% confidence intervals for cost-effectiveness ratios is the nonparametric bootstrap method. 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 (best) to highest (worst); 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.

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.

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.

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. 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.

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, 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.

Key Points

- Ideally, in economic evaluations of clinical trials, the economic study is integrated into the clinical protocol and the economic data are collected as part of a unified case report form for both clinical and economic variables.
- In general, programs that cost more and are more effective (and perhaps even some programs that cost less and have reduced clinical outcomes) should be adopted if both their cost-effectiveness and incremental cost-effectiveness ratios fall within an acceptable range.
- To address concerns about their reliability and validity, all economic evaluations should include sensitivity analyses of variables that have a significant effect on the study's conclusions but for which values are uncertain.
- Results of an economic analysis are just one component of the decision-making process regarding the adoption of an intervention; social, legal, political, and ethical issues, among others, are also important.

SUGGESTED FURTHER READINGS

- Bombardier C, Eisenberg J. Looking into the crystal ball: can we estimate the lifetime cost of rheumatoid arthritis? *J Rheumatol* 1985; **12**: 201–4.
- Brown M, Glick HA, Harrell F, Herndon J, McCabe M, Moynour 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.
- 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.
- Detsky AS, Naglie IG. A clinician's guide to cost effectiveness analysis. *Ann Intern Med* 1990; **113**: 147–54.
- Drummond MF, Stoddart GL, Torrance GW. *Methods for the Evaluation of Health Care Programs*. New York: Oxford Medical Publications, 1987.
- Eddy DM. Cost-effectiveness analysis: is it up to the task? *JAMA* 1992; **267**: 3342–8.
- Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
- Eisenberg JM. From clinical epidemiology to clinical economics. *J Gen Intern Med* 1988; **3**: 299–300.
- Eisenberg JM. Clinical economics: a guide to the economic analysis of clinical practices. *JAMA* 1989; **262**: 2879–86.
- Finkler SA. The distinction between cost and charges. *Ann Intern Med* 1982; **96**: 102–9.
- Granneman TW, Brown RS, Pauly MV. Estimating hospital costs. *J Health Econ* 1986; **5**: 107–27.
- Hsiao WC, Braun P, 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.
- Laska EM, Meisner M, Seigel C. Power and sample size in cost-effectiveness analysis. *Med Decis Making* 1999; **19**: 339–43.
- 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.
- 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.
- 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.
- 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.
- Siegel E, Russell LB, Weinstein C, Gold MR. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
- Willan AR. Analysis, sample size, and power for estimating incremental net health benefit from clinical trial data. *Control Clin Trials* 2001; **22**: 228–37.
- Willke RJ, Glick HA, Polsky D, Schulman K. Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Econ* 1998; **7**: 481–93.

Using Quality-of-Life Measurements in Pharmacoepidemiologic Research

The following individuals contributed to editing sections of this chapter:

HOLGER SCHÜNEMANN,¹ GORDON H. GUYATT,^{2,3} and ROMAN JAESCHKE³

¹Department of Medicine, University at Buffalo, State University of New York, USA, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada and Division of Clinical Research Development and Information Translation/INFORMA, Italian National Cancer Institute, Rome, Italy; ²Department of Medicine and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; ³Department of Medicine, St. Joseph's Hospital, and Department of Medicine, McMaster University, 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, which 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, 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. 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).

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 process. It follows 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.

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 (1986) studied the effect of three established antihypertensive drugs—captopril, methyldopa, and propranolol—on quality-of-life, long after their introduction in clinical practice. 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 might 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. Investigators must define as precisely as possible the aspects of HRQL in which they are interested (e.g., a specific domain or general HRQL).

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. 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.

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. 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) and *construct 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 the 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 in HRQL among patients. In this context, the way of quantifying the signal-to-noise ratio is called *reliability*. If the variability in scores among subjects (the signal) is much greater than the variability within 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.

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 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. Investigators have suggested other measurements of responsiveness that all rely on some way of relating signal to noise.

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 the HRQL score by 3 points relative to control, what are we to conclude? Is the treatment effect very large, warranting widespread dissemination, or is it trivial, suggesting that the new treatment should be abandoned? This question highlights the importance of being able to interpret the results of our HRQL questionnaire scores.

Researchers have developed a number of strategies to interpret HRQL scores. 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.

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. 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 1-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; symptoms, emotional function, and activity limitations in both adult and child asthma patients, and parents of child asthma patients; and symptoms, emotional function, and activity limitations

in adults with rhinoconjunctivitis. Similar observations may be derived from reports of other investigators.

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. 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.

Another anchor-based approach uses HRQL instruments for which investigators have established the minimal important difference (MID, see Case Example 23.1). 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. 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 (e.g., a correlation coefficient of 0.5).

CASE EXAMPLE 23.1: **INTERPRETING RESULTS OF CLINICAL** **TRIALS FOCUSING ON HRQL OUTCOMES**

Background

- Clinical studies in patients with chronic obstructive pulmonary disease (COPD) rely on measuring health-related quality-of-life, because physiological measures do not correlate highly with how patients feel. Investigators used a health-related quality-of-life instrument to measure whether patients with COPD had improved dyspnea after a drug intervention. They measured a mean change of 0.8 on the 7-point scale of the Chronic Respiratory Questionnaire (CRQ) dyspnea domain in the intervention group and no change in the control group.

Question

- What have researchers done to help readers and clinicians answer the question “What does a change of 0.8 on the 7-point scale of the CRQ mean”?

Approach

- Investigators have used several techniques to investigate the interpretability of the CRQ and other health-related quality-of-life instruments. They have determined what constitutes the minimal important difference (MID) using comparisons with global ratings of change, with other instrument for which interpretability is known, and distribution-based statistical methods.

Results

- A change of 0.5 or more on the CRQ dyspnea domain indicates that patients, on average, have experienced an important change. The majority of patients in this study who received the intervention experienced an important improvement.

Strengths

- Multiple methods have shown a similar magnitude for the MID on the CRQ.
- Clinicians can interpret changes on health-related quality-of-life instruments if they know what constitutes an important change.

Limitation

- There is no gold standard for measuring health-related quality-of-life or for measuring interpretability.

Summary Points

- Health-related quality-of-life is a key outcome in chronic disease such as COPD.
- Clinicians need to be able to interpret results of health-related quality-of-life outcomes.
- Several techniques exist that facilitate interpretation of health-related quality-of-life changes.

Investigators proposed distribution-based methods to determine the 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. 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. 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. The problem with 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 that they learn to understand other HRQL scores.

Cohen (1988) 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. 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?

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 quality-of-life measure.

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 if they had 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% of the treated patients did not improve at all. Summary measures

would yield a mean change of 0.25 in the treated patients, and a median change of zero in this group, suggesting minimal effect. Yet, 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 25–50 patients, and often as many as 100, to prevent a single adverse event. 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. In a crossover randomized trial in asthmatic patients comparing the short-acting inhaled β -agonist salbutamol to the long-acting inhaled β -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 of 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.

This discussion emphasizes that interpreting the results of HRQL measurement in pharmacoepidemiology 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

We have also suggested a taxonomy based on the domains of HRQL which an instrument attempts to cover. According to this taxonomy, an HRQL instrument may be categorized as generic or specific. *Generic instruments* 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. In contrast, *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 Sickness Impact Profile (SIP), contains twelve “categories” which can be aggregated into two dimensions and five independent categories, and also into a single overall score. Increasingly, a collection of related instruments from the Medical Outcomes Study have become the most popular and widely used generic instruments. Particularly popular is one version that includes 36 items, the SF-36. The SF-36 is available in over 40 languages, and normal values for the general population in many countries are available.

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. 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 or particular disease. 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. We will return to this issue in the section on specific instruments.

Utility Measurement

Economic and decision theory provides the underlying basis for *utility measures*. The key elements of a utility instrument are, first, that it is preference-based, 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, the Health Utilities Index, and the Euroqol (EQ5).

The second approach is to ask patients to make a single rating that takes into account all aspects of their quality-of-life. 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.” 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 22). 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. Cost per QALY may be compared and provides a basis for allocation of scarce resources among different health care programs.

However, utility measurement also has limitations. Utilities can vary depending on how they are obtained, raising questions of the validity of any single measurement. Utility measurement does not allow the investigator to determine which aspects of HRQL are responsible for changes in utility and 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 (e.g., for chronic lung disease,

for rheumatoid arthritis, for cardiovascular diseases, for endocrine problems), a population of patients (e.g., the frail elderly, who are afflicted with a wide variety of different diseases), a certain function (e.g., emotional or sexual function), or 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.

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 (e.g., dyspnea) is related to the magnitude of physiological abnormality (e.g., exercise capacity). 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.

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. 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. However, use of multiple instruments requires caution about how to interpret results if they differ between instruments and because of multiple testing of statistical hypotheses.

THE FUTURE

The considerations we have raised suggest a step-by-step approach to addressing issues of HRQL in pharmacoepidemiology studies. 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 that 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 that 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.

Key Points

- Health-related quality-of-life has become an established outcome measure in clinical research.
- The *validity* of an instrument refers to its ability to measure what it is intended to measure, while responsiveness determines the signal-to-noise ratio and refers to an instrument's ability to detect change.
- Discriminative instruments aim to measure differences among patients at one point in time, while evaluative instruments aim to measure changes over time.
- Interpretability is a key measurement property of an instrument and relates to the extent to which one can understand the magnitude of any differences between treatments that a study demonstrates.
- Disease-specific instruments offer greater responsiveness compared with generic instruments.

SUGGESTED FURTHER READINGS

- 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.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
- Croog S, Levine S, Testa M. The effects of antihypertensive therapy on the quality-of-life. *N Engl J Med* 1986; **314**: 1657–64.

- Gill T, Feinstein A. A critical appraisal of the quality of quality-of-life measurements. *JAMA* 1994; **272**: 619–26.
- 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.
- 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.
- Guyatt G, Feeny D, Patrick D. Measuring health-related quality-of-life: basic sciences review. *Ann Intern Med* 1993; **70**: 225–30.
- Guyatt G, Juniper E, Walter S, Griffith L, Goldstein R. Interpreting treatment effects in randomized trials. *BMJ* 1998; **316**: 690–3.
- Jaeschke R, Guyatt G, Keller J, 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.
- Jaeschke R, Singer J, Guyatt G. Using quality-of-life measures to elucidate mechanism of action. *Can Med Assoc J* 1991; **144**: 35–9.
- 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.
- Schünemann HJ, Guyatt GH. Commentary—Goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res* 2005; **40**: 593–7.
- Ware J, Sherbourne C. The MOS 36-item short-form health survey (SF-36). *Med Care* 1992; **30**: 473–83.
- 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.

The Use of Meta-analysis in Pharmacoepidemiology

The following individuals contributed to editing sections of this chapter:

CARIN J. KIM¹ and JESSE A. BERLIN²

¹University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; ²Johnson & Johnson Pharmaceutical Research and Development.

INTRODUCTION

DEFINITIONS

Meta-analysis has been defined as “the statistical analysis of a collection of analytic results for the purpose of integrating the findings.” It 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. 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, 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. Meta-analysis of randomized clinical trials has found such wide acceptance that an international organization, the Cochrane Collaboration, with an associated electronic library, has been built around the performance and updating of meta-analyses of trials (<http://www.cochrane.org>). A similar structure has developed in the social sciences, in the form of the Campbell Collaboration.

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. 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. 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 rigorous literature reviews; similar measures from comparable studies are tabulated systematically and the effect measures are combined when appropriate.

The most popular conception of meta-analysis is the summary of a group of randomized clinical trials dealing with a particular therapy for a particular disease. Typically, a meta-analysis would present an overall measure of the efficacy of treatment, e.g., a summary odds ratio. Summary measures may be presented for different subsets of trials

involving specific types of patients, e.g., 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 attempt to uncover the sources of that variability.

More recently, meta-analyses of nonexperimental epidemiologic studies have been performed. In general, the meta-analyses of nonexperimental studies and the associated methodologic articles tend to focus more on the exploration of reasons for disagreement among the results of prior studies, including the possibility of bias. This chapter summarizes many of the major conceptual and methodologic issues surrounding meta-analysis and offers the authors' views about possible avenues for future research in this field.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGIC RESEARCH

The reasons why a pharmacoepidemiologist might be interested in conducting a meta-analysis include the study of uncommon adverse outcomes of therapies free of the confounding and bias of nonexperimental 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. Examples throughout this chapter will serve to illustrate these applications.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGIC 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.

QUALITY OF THE ORIGINAL STUDIES

Meta-analysis, by combining a group of poorly done studies, can produce a spuriously precise summary result (i.e., an incorrect measure of effect that has narrow confidence

intervals) that should not be used as a basis for formulating clinical or policy strategies. However, if the quality assessment is influenced by the direction or magnitude of the findings, bias in the meta-analytic process could result. It should also be recognized that those clinical and policy strategies need to be based on something, so the question becomes whether relying on a collection of poorly done studies is better or worse than relying on the consensus of "experts," whose opinions may or may not have evidentiary support.

COMBINABILITY OF STUDIES

Clearly, no one would suggest combining studies that are so diverse that a summary would be nonsensical. Should studies with different patient populations be combined? How different can those populations be before it becomes unacceptable to combine the studies? Should nonrandomized studies be combined with randomized studies? Should nonrandomized studies ever be used in a meta-analysis? There is unlikely to be unanimous agreement on answers to these questions.

PUBLICATION BIAS

Publication bias occurs when study results are not published, or their publication is delayed because of the results. The usual concern is that statistically significant results are published more easily than nonsignificant results. Publication bias is a potentially serious limitation to any meta-analysis because unpublished data can represent a large proportion of all available data.

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. In particular, bias can result from the process of selecting and rejecting the studies to include in the meta-analysis (as discussed in more detail later in this chapter). For example, 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 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) showed only a trend. These disagreements were not easily explainable.

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. In the second part of this section, specific solutions to the methodologic issues raised in the previous section are presented. Finally, case study examples of applications that should be of interest to pharmacoepidemiologists will be presented.

STEPS INVOLVED IN PERFORMING A META-ANALYSIS (TABLE 24.1)

Define the Purpose

It is particularly important to define precisely, and in advance of its conduct, the primary and secondary objectives of a meta-analysis.

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. 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, whereas use of too many restrictions can result in missing a substantial number of relevant publications. The help of a trained professional librarian, a review of the relevant reference sections of retrieved publications, and hand searches of relevant journals are 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. Practical considerations may, of course, force changes in the inclusion criteria. For example, one might find no randomized studies of a particular new indication for an existing therapeutic agent, thus forcing consideration of nonrandomized studies.

In establishing inclusion/exclusion criteria, one is also necessarily defining the question being addressed by the meta-analysis. The use of broad entry criteria permits the examination of the effects of research design on outcome (e.g., do randomized and nonrandomized studies tend to show different effects of therapy?) or the exploration of subgroup effects. 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 not 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. Another important note is that studies may often generate more than one published paper. It is essential that only one report on the same patients be accepted into the meta-analysis. 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. Although there is no general rule that we can recommend for choosing one publication over others, reporting clearly the reasons for choosing one publication over others is absolutely essential.

Collect the Data

When the relevant studies have been identified and retrieved, the important information regarding study design and outcome needs to be extracted and recorded on data abstraction forms. Careful specification in the protocol for the meta-analysis of the design features and patient characteristics that will be of interest may help to avoid over- or under-collecting information.

Many articles on "how to do a meta-analysis" recommend that the meta-analyst assess the quality of the studies being considered in a meta-analysis using formal quality scoring systems. However, limitations of formal quality scoring systems (which tend to measure and weight a combination of completeness of reporting and factors that might relate to the potential for bias) have led to more recent recommendations that specific aspects of study design or conduct be explored as predictors of study findings. For example, one might explore whether studies with adequate concealment of randomization produce systematically different results from studies with inadequate concealment.

Table 24.1. General steps involved in conducting a meta-analysis

-
- (1) Define the purpose
 - (2) Perform the literature search
 - (3) Establish inclusion/exclusion criteria
 - (4) Collect the data
 - (5) Perform statistical analyses
 - (6) Formulate conclusions and recommendations
-

Perform Statistical Analyses

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. 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. Other approaches include inverse-variance weighted averages of study-specific estimates of multivariate-adjusted relative risks and exact stratified odds ratios. One popular method, sometimes called the “one-step” method, is similar to the Mantel–Haenszel method but has been shown to be biased under some circumstances. However, in a simulation study, Deeks and colleagues showed that in situations where there are rare events, and consequently frequent zero cells in contingency tables, the one-step method tended to perform better than other alternatives, including Mantel–Haenszel and exact methods.

A basic principle in most analytic approaches is that the comparisons between treated (exposed) and untreated (unexposed) patients are made within a study prior to combination across studies. In any meta-analysis, the possible existence of heterogeneity among study designs and results should be examined, and may warrant a set of exploratory analyses designed to investigate the sources of that heterogeneity. Depending on the results of such exploratory analyses, 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. Bayesian approaches to meta-analysis, which provide a more flexible, but more complex, modeling structure, have also been proposed. Sophisticated models, though, are *not* a substitute for a thorough exploration of reasons for heterogeneity of results across studies. 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. In addition to stratification, regression methods such as weighted least-squares linear regression could be used to explore sources of heterogeneity. These might

be important when various components of study design are correlated with each other, acting as potential confounders. In recent years, increasingly sophisticated (and complex) approaches to the statistical modeling of heterogeneity have been proposed, e.g., different forms of weighted normal errors regression and random effects logistic regression. Helpful overviews of statistical methods for meta-analysis are provided in a number of sources, including a recent text (Sutton, Abrams *et al.*, 2000) and several review articles. Tutorials on basic and more advanced meta-regression methods are also available.

Formulate Conclusions and Recommendations

Conclusions of a meta-analysis should be clearly summarized, with appropriate interpretation of the strengths and weaknesses of the meta-analysis. Authors should clearly describe the population(s) to which the results apply, and how definitive the results are, and should outline the areas that need future research.

APPROACHES TO SELECTED METHODOLOGIC PROBLEMS IN META-ANALYSIS

Combinability of Results from Diverse Studies: Heterogeneity is Your Friend

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. Conversely, diversity of designs and results may provide an opportunity to understand the factors that modify the effectiveness (or toxicity) of a drug. A helpful distinction to keep in mind when considering sources of variability of results is between aspects of study design, which represent “artifact,” and features of study populations or treatments, which represent potential biological reasons for variable results.

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.

As an example of the type of analysis that could be used to investigate study design issues, Hennessy and colleagues (2001) performed a meta-analysis of nonexperimental studies comparing third-generation oral contraceptives (those containing gestodene and desogestrel) to second-generation

pills (those containing levonorgestrel) with respect to the risk of venous thromboembolic events. 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, suggesting that depletion of susceptibles did not explain the association between third-generation pills and thromboembolic events. 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. For example, Morgenstern and colleagues (1987) found that the association between neuroleptic medication and tardive dyskinesia was stronger in studies conducted in the United States than in studies conducted elsewhere. They used regression methods to show that this association was not simply the product of confounding by other study design features. The authors suggest that the US study samples may have had a higher baseline frequency of unmeasured factors (e.g., affective disorders such as schizophrenia) than the exposed groups in other countries. As with any exploratory analysis, due caution must be exercised in the interpretation of such *a posteriori* hypotheses, even though they may be based on very sound biological reasoning.

Publication Bias

As discussed above, when the primary source of data for a meta-analysis is published data, there is usually a danger that the published studies represent a biased subset of all the studies that have been done. In general, it is more likely that studies with statistically significant findings will be published than studies with no significant findings. There are graphical methods to help determine this potential for

publication bias. A practical technique is the "funnel plot." 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. An asymmetry or a bite-out of the funnel shape will indicate possible existence of publication bias. Sterne and Egger (2001) point out that publication bias is just one possible explanation of the funnel asymmetry so that the funnel plot should be seen as estimating "small study effects" (i.e., smaller studies tend to show larger treatment effects in a meta-analysis) rather than necessarily publication bias.

Several mathematical approaches to the problem of publication bias have been proposed. An early method, first described by Rosenthal, 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 zero that would be required to exist, in order for the combined *Z*-score (published + unpublished studies) to become nonsignificant. Because this method focuses only on *Z*-statistics, and ignores the estimation of effects (e.g., odds ratios), it is of limited utility. That is, the fail-safe *N* approach focuses only on the statistical significance of the combined result and does not help to provide an overall estimate of the effect that is "adjusted" for publication bias.

A proposed solution to the problem of publication bias is the use of prospective registration of studies at their inception, prior to the availability of results. Going one step further, several prospective meta-analyses are either being planned or have been conducted. 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.

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. 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 vastly increased statistical power to investigate adverse events. In fact, since 1982, the safety evaluation of drugs in the US has included pooled analyses.

The assessment of the excess risk of gastrointestinal side effects associated with NSAIDs provides an excellent example of a situation in which meta-analysis has been helpful. Case examples of the meta-analytic approach to this problem will be reviewed here.

**CASE EXAMPLE 24.1: THE APPLICATION
OF META-ANALYSIS IN ASSESSING
ADVERSE EVENTS: NSAIDs STUDIES BY
VARIOUS AUTHORS INCLUDING CHALMERS
ET AL. (1988) AND DEEKS ET AL. (1998)**

Background

- Cohort or case-control studies investigating nonsteroidal anti-inflammatory drugs (NSAIDs) as risk factors for gastrointestinal side effects are subject to too many potential biases.

Question

- Can combining data from randomized trials of NSAIDs alleviate the problem of bias in observational studies and would doing so allow us to assess the risk of uncommon adverse events?

Approach

- Perform meta-analysis of randomized trials with certain inclusion/exclusion criteria.

Results

- A generally low risk of gastrointestinal side effects.
- Duration-response effect observed (studies with longer duration showed harmful effect of NSAIDs).

Strengths

- Bias can be alleviated by combining different studies.
- Even with the frequent zero cells, analysis can be carried out.

Limitations

- Many studies did not mention side effects and the authors assumed these studies to have zero side effects.

- Too many zero cells that do not contribute to estimates of an overall summary odds ratio.

Summary Points

- Meta-analysis can assess risk of adverse events as seen with the increasing risk with increasing use of duration of NSAIDs.
- One can use risk differences to include studies with no events in the calculations.

Chalmers and colleagues (1988) examined data from randomized trials of NSAIDs (see Case Example 24.1). They argued that the typical epidemiologic approaches to investigating NSAIDs as risk factors for gastrointestinal side effects, i.e., cohort or case-control studies, are subject to too many potential biases. Randomized trials, on the other hand, would provide internally valid comparisons of NSAID users to nonusers. Presumably, although not stated explicitly, the combination of results from numerous studies, with varied entry and exclusion criteria, would alleviate the problem of the potential lack of generalizability from patients enrolled in a particular trial. The pooling of results from numerous studies would permit the assessment of rare events.

The authors performed a meta-analysis of randomized trials, limited to those trials in which the anti-inflammatory drug was compared with a placebo, no drug, or a drug with no anti-inflammatory property. Photocopies of the "Methods" sections of 525 potentially relevant studies (blinded as to author, journal, and time and place of study, as well as all allusions to results) were read by two independent observers who determined inclusion suitability according to the above criteria.

As a methodologic aside, the authors examined inter-reader disagreements. Overall, a disagreement rate of 19% was observed for the final decision on inclusion or exclusion of studies. These disagreements were resolved in conference.

There were 100 randomized trials of non-aspirin NSAIDs included in the final analysis, containing 123 comparisons with a no-treatment control group, which usually received a placebo. A total of 12 853 patients were included in these trials, with a mean duration of treatment of about 67 days (median 21 days) and a mean age of 46 years. For the sake of brevity, the aspirin trials will not be discussed here. The data revealed a generally low risk of gastrointestinal side effects. For example, only two patients were reported with

proven ulcers out of 6460 treated patients, with none in the controls. In the 10 studies explicitly mentioning gross upper-gastrointestinal hemorrhage, the risk was 8/1103 (0.73%) in the control patients and 24/1157 (2.1%) in treated patients, giving a crude relative risk of 2.8. The length of follow-up for these 10 studies was not specifically mentioned by the authors of the meta-analysis. However, the analysis of duration of therapy showed that duration was longer for studies showing a harmful effect of NSAIDs (geometric mean = 81 days) than for studies showing no effect of NSAIDs (geometric mean = 25 days) for the gross hemorrhage endpoint, consistent with a duration–response effect.

This meta-analysis was faced with some interesting statistical and other methodologic questions. There were numerous studies that did not explicitly mention side effects in general or did not mention particular side effects, even though others were mentioned. The authors chose to do a kind of sensitivity analysis by analyzing all studies, assuming that the risk of an unreported side effect was zero, and separately analyzing results from only those studies explicitly mentioning a particular side effect.

Another issue was the extensive number of studies with no occurrences of a particular endpoint in either the treated or the control group. The usual pooling procedures, e.g., the Mantel–Haenszel procedure, essentially ignore such studies, since they contribute no information, under one interpretation, concerning the common odds ratio. On the other hand, if over 90 of 100 separate trials report no proven ulcers in either the treated or the control groups, then another interpretation of those results is that the risk of an ulcer is fairly low. Chalmers and colleagues chose to work with risk differences to address this issue, allowing studies with no events in either group to enter the calculations. This is the type of situation considered by Deeks and colleagues (1998), whose results suggest that the one-step method would have been the most appropriate for these studies with frequent occurrence of zero cells.

In perhaps the most comprehensive and clinically useful of the systematic reviews in the area of NSAIDs side effects, Henry and colleagues (1996) addressed the issue of *comparative* relative risks of serious gastrointestinal complications with individual NSAIDs (see Case Example 24.2). Their stated motivation for this approach was that one strategy for reducing NSAID toxicity in populations would be to choose, as first line therapy, a drug and dose with a comparatively low risk of gastrointestinal side effects.

CASE EXAMPLE 24.2: **THE APPLICATION OF** **META-ANALYSIS IN ASSESSING** **ADVERSE EVENTS**

Background

- It would be desirable to choose a drug and dose with a comparatively low risk of gastrointestinal side effects to reduce NSAID toxicity in populations.

Questions

- Can one compare relative risks of serious gastrointestinal complications across individual NSAIDs?
- To what extent could differences in toxicity be related to different doses (or “dose equivalents”)?

Approach

- Perform meta-analysis to examine relative risks for particular NSAIDs.
- Weighted ranking system that incorporates study size into the weights implemented.
- Comparisons among NSAIDs were indirect, as very few head-to-head studies were available.

Results

- Relative risks for all NSAIDs were significantly greater than 1.0.
- Weighted analysis showed ibuprofen with the lowest and ketoprofen with the highest risks.
- High dose ibuprofen also was associated with elevated relative risk.

Strengths

- Systematic review of all relevant data.
- Can apply weights to studies.

Limitations

- Indirect comparisons to ibuprofen depend on assumptions about similarities of populations studied in different studies.
- Analysis limited to studies available (only studies that compared to ibuprofen as reference were selected).

(Continued)

Summary Points

- By using meta-analysis, one can systematically review NSAID side effects for specific drugs.
- Studies can be weighted according to their size.
- Systematic review may provide us with helpful information in clinical decision making.

The authors used meta-analytic methods to examine the range of relative risks for particular NSAIDs and explore the extent to which differences in toxicity could be related to different doses, or to different susceptibility among patients receiving the various drugs. To do this, they identified case-control or cohort studies of relationships between use of specific NSAIDs in the community and development of serious peptic ulcer complications requiring hospital admission. In estimating pooled relative risks, analyses were restricted to studies that compared another drug with ibuprofen as the reference. They used unadjusted relative risks based on 2×2 tables in the pooling.

The authors found 12 studies examining 14 NSAIDs, including two unpublished reports. Eleven of the studies were case-control studies. The estimated relative risks for specific drugs versus ibuprofen ranged from 1.6 (CI 1.0, 2.5) for fenoprofen, to 9.2 (CI 4.0, 21) for azapropazone. All of the relative risks were significantly greater than 1.0. Using a weighted ranking system, which incorporated study size into the weights, the authors found that ibuprofen had the lowest rank (least toxicity), followed by diclofenac. Aspirin and naproxen had intermediate risks, while azapropazone, tolmetin, and ketoprofen had the highest risks. High dose ibuprofen (i.e., greater than 1600 mg daily) was also associated with an elevated relative risk.

It is important to keep in mind that the conclusions reached by Henry *et al.* were based on indirect comparisons of the various drugs with ibuprofen. They claimed to find little evidence that the relative rankings were due to confounding by patient susceptibility. Despite any shortcomings of their approach, as the authors point out, clinical and regulatory decisions have to be made on some type of scientific basis, and these are the only data available. Risks need to be weighed against benefit, and the authors highlight the known variability across patients in clinical response to particular drugs. Thus, it seems that this systematic review provided useful information for clinical decision making.

In qualitative reviews of the literature on the gastrointestinal side effects of NSAIDs, Taragin *et al.* (1990) and Carson and Strom (1992) point out differences in study designs that

could lead to differences in results. For example, bleeding could be defined as all bleeding, fatal bleeding, bleeding requiring hospitalization, bleeding requiring transfusions, or bleeding requiring surgery. Several procedures exist for the detection of gastrointestinal bleeding. The clinical relevance of the different methods is sometimes unclear. Case-control studies may show higher odds ratios because of the likelihood of recall bias; patients with bleeding requiring hospitalization might be more likely to recall NSAID use than controls, particularly if probing by interviewers or by health care providers prior to interview is more extensive for cases than for controls. This possibility is supported by the data from the Gabriel *et al.* (1991) meta-analysis. Cohort studies based on claims data, such as that conducted by Carson and colleagues (1987), sometimes use unvalidated outcomes. To the extent that false events may be documented for both the exposed and unexposed cohorts, the relative risk observed in such studies would show less of an effect of exposure. Of course, these cohort studies may exaggerate the apparent effect of exposure if spurious diagnoses of gastrointestinal events are more likely to occur when a patient has a history of NSAID use. Further variability may be generated among study results by the inclusion of many different kinds of NSAIDs, some of which may have more potential to cause gastrointestinal side effects than others.

Thus, another benefit of meta-analysis is the ability to examine findings according to study characteristics and study design, leading to hypotheses about subgroups or particular therapies of special interest and suggestions for the design of subsequent studies. Meta-analysis can quantify differences related to study design that the traditional review can only observe in qualitative terms.

There are numerous other examples of the application of meta-analysis to the evaluation of adverse effects of pharmaceutical therapies.

New Indications for Existing Therapies

Meta-analysis has also been used to assess the effectiveness of existing therapies for new indications. As an example, the efficacy of antilymphocyte antibodies in the perioperative period of cadaveric kidney transplantation (induction therapy) had not, until recently, been conclusively demonstrated. Individual studies, both randomized and nonexperimental, had failed individually to show a significant benefit of induction therapy with respect to allograft survival. Szczech and colleagues (1997) performed a meta-analysis of the published data from the randomized trials of induction therapy in adults receiving cadaveric renal transplants. That analysis, using survival analytic methods

on the group-level (published) data, showed a statistically significant 31% lower rate of allograft failure at two years in patients receiving induction therapy.

In a subsequent analysis of the individual patient data from five of the seven randomized trials of induction therapy, Szczech and colleagues (1998) examined the effect of induction therapy beyond two years and in subgroups of patients with risk factors for early allograft failure. The subgroup analyses are examined in the next section. The five studies included in the individual patient analyses yielded results for the two-year analysis virtually identical to those obtained from the full set of seven studies using the published data, i.e., a relative rate of 0.69 favoring induction therapy. When extended to five years, the rate of allograft survival was 69.0% in patients receiving induction therapy and 64.4% in those not receiving induction therapy ($p = 0.13$). Thus, the overall benefit demonstrated at two years was smaller and no longer significant at five years.

Differential Effects among Subgroups of Patients

In the analysis of individual patient data by Szczech and colleagues (1998), the authors were able to examine the specific effects of induction therapy in subgroups of patients at high risk for allograft failure. Before proceeding to analyses *within* particular subgroups, the authors tested statistical interactions between each of the relevant patient characteristics and induction therapy. One of the patient characteristics of interest was the panel reactive antibody (PRA) level, an indicator of immune system presensitization. Patients with PRA levels less than 20% were considered unsensitized, while those with PRA of 20% or higher were considered presensitized. At two years, the effect of induction therapy differed in presensitized and unsensitized patients ($p = 0.03$ for interaction). The rate ratio at two years was 0.12 (CI 0.03–0.44, $p = 0.002$) in presensitized patients (85 patients with 15 failures) and 0.74 (CI 0.50–1.09, $p = 0.13$) in unsensitized patients (511 patients with 100 failures). This interaction was still significant at five years ($p = 0.009$ for interaction), with a rate ratio of 0.20 (CI 0.09–0.47, $p < 0.001$) in presensitized patients (85 patients with 33 failures) and 0.97 (CI 0.71–1.32, $p > 0.2$) in unsensitized patients (510 patients with 163 failures). The authors found no other significant interactions between induction therapy and any other variable.

Several advantages of meta-analysis, and particularly individual-patient analyses, are demonstrated by this example. The improved precision provided by large numbers of patients is an important benefit. Having individual-level data allowed an analysis that could go beyond the simple, unadjusted analyses to which most meta-analyses of published

data are limited. The availability of patient characteristics permitted not only *adjustment* for those characteristics, but also examination of subgroup effects in larger numbers of patients than would typically be included in a single trial. Although one might wish to confirm these subgroup results in an independent data set, the patient-level analyses strongly suggest that induction therapy is effective in the 14% of patients who are presensitized. If confirmed, these results could mean that induction therapy could be targeted to the group in which it is highly effective, while avoiding needless treatment and potential toxicity in other patients.

Selection from among Several Alternative Therapies

In a meta-analysis of therapies for the prevention of supraventricular arrhythmias after coronary bypass surgery, Andrews *et al.* (1991) looked separately at verapamil, digoxin, and β -adrenoceptor blockers as prophylactic agents. Only randomized trials were included. Neither digoxin nor verapamil reduced the risk of supraventricular arrhythmias after coronary artery bypass surgery (digoxin: OR = 0.97, CI = 0.62, 1.49; verapamil: OR = 0.91, CI = 0.57, 1.46). The risk of a supraventricular arrhythmia in patients treated with beta-blockers was dramatically reduced (OR = 0.28, CI = 0.21, 0.36), although significant heterogeneity among the study results was present. The authors explored this heterogeneity by examining separately studies of different beta-blockers, and by summarizing separately preoperative and postoperative treatment. While these separate summaries suggested varying degrees of heterogeneity within subgroups of studies, all of the summaries showed statistically significant benefits of beta-blockers. The authors drew no firm conclusions from the subgroup analyses other than to suggest directions for future research.

Saving Time and Money if You Believe a Meta-analysis

One of the potential benefits of meta-analysis is the possibility of shortening 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 (2002) report the use of meta-analysis of ECG data from several clinical pharmacology studies for two 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,” i.e., performing a new meta-analysis each time the results of a new clinical trial are published. Antman *et al.* (1992) applied this 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.

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. At the very 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. Another consideration is that estimates of treatment effect may not be stable over time, perhaps due to changing clinical environments. 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 nonexperimental studies, including cohort and case-control studies and nonrandomized Phase II clinical trials. The results of these studies are not always confirmed by subsequent, properly designed randomized trials. For example, consider the case of β -carotene in the prevention of cancer. A series of observational studies examined the relation between dietary intake of foods rich in β -carotene and the risk of lung cancer. Overall, they showed a relatively consistent association between diets rich in β -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.

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 availability of

individual-level data. From the above examples, it becomes increasingly apparent that the pursuit of questions about subgroups of patients is often an informative and important element of a well-conducted meta-analysis, at least for certain therapies. One should certainly exercise due caution in the interpretation of subgroup effects, emphasizing those that are specified *a priori* with biological justification. By assembling large numbers of patients, meta-analysis can at least begin to address the problems related to statistical instability of subgroup effects. It is too often the case, however, that results are not reported separately for subgroups of patients. Typically, some trials will exclude particular patients while others will not exclude them. At the level of grouped data from published reports, one is faced with analyzing the two groups of studies separately as the only way of addressing the subgroup question, or using so-called “meta-regression” techniques to explain heterogeneity in terms of particular study-level covariate(s). For example, one might perform a regression of treatment effect on the percent male subjects at the study level, to try to address whether treatment effect differs between men and women.

In practice, an important question is whether results for subgroups obtained using group-level data are consistent with what one would find using patient-level data. Increasing evidence is accumulating that such consistency is not the rule. In a methodologic paper extending the work on induction therapy described above, Berlin and colleagues (2002) found that the conclusion that induction therapy in renal transplant patients is limited to those with elevated PRA was not found consistently in the group-level data. Lambert and colleagues (2002), in a simulation study, point out that the group-level analyses often have low statistical power, even in the absence of systematic bias in the group-level comparisons. Thus, it seems there is a trade-off between the resource-intensive individual patient analysis against the less expensive but potentially invalid group-level analysis. Again, the need to facilitate availability of individual patient data is apparent and has been recognized by the US government. The National Institutes of Health (NIH) in the US have developed a policy on data sharing of final research data from NIH-supported studies for use by other researchers. Investigators submitting an NIH application will be required to include a plan for data sharing or to state why data sharing is not possible. The NIH will support such sharing, financially, either in project budgets or through administrative supplements to grants. (For more details, see <http://grants2.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.)

In the development of cumulative meta-analysis, some of the most important issues will be philosophical ones. Some

of the same issues apply to the approval process for new drugs. How much evidence is required before a therapy can be accepted as efficacious? Should we require the existence of a certain minimum number of trials showing a statistically significant benefit of a therapy? In this context, it is worth noting that several empirical studies have examined discrepancies between large trials and meta-analyses of the same therapies. The assumption made by some of the authors of these studies, that larger studies are necessarily better studies, may not be valid. Replication of a finding by independent studies must certainly be a key element to establishing efficacy, as compared with a single trial. Large trials may also be poorly designed. When there is little or no heterogeneity of results among trials, and the likelihood of serious publication bias is minimal, 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, but such a conclusion might be conditional on the type of patient or other factors.

Similarly, the technique of cumulative meta-analysis could be applied to the analysis of adverse events. As nonexperimental studies of adverse effects are completed, the same approach could be applied. The likelihood seems to be, however, that such meta-analyses would be faced with much more serious issues of heterogeneity of findings than meta-analyses of randomized studies typically have to confront. The acceptance of meta-analytic results in this context might be extremely slow.

There has been growing attention paid to the potential for indirect comparisons to contribute evidence regarding the efficacy and safety of therapies. Alternative treatments for the same indication are often available, but have not always been evaluated in head-to-head randomized comparisons. Even when there are direct comparisons available, these may constitute a small portion of the available evidence. Several papers have presented statistical methods for performing indirect comparisons. The basic principle of these methods is most clearly discussed by Bucher and colleagues (1997). If drug A is compared to placebo in one randomized trial using an odds ratio for a dichotomous outcome, and drug B is similarly compared to placebo in another randomized trial, the comparison of A to B is achieved essentially by dividing the odds ratios, in effect "subtracting out" the placebo. The assumption needed for these indirect comparisons to be valid is that there is no modification of the treatment effect by patient characteristics that may differ in their distribution

between the two trials (even if that factor is balanced within each trial). So, for example, if treatments A and B in the hypothetical trials above were both effective in men but not in women, the trial of drug A had 20% men in each of the active and placebo arms, and the trial of drug B had 50% men in each arm, drug B could appear to more effective than drug A even if the two drugs were truly equally effective, simply because of the imbalance *between* the trials in the sex distribution of the patients. Song and colleagues (2000) reviewed 44 published meta-analyses in which both direct and indirect comparisons had been made, and found that adjusted indirect comparisons often, but not always, agree with the corresponding direct comparisons. In 3 of the 44 comparisons, a significant discrepancy between the adjusted indirect and the direct comparisons was found.

The concept of prospective meta-analysis also merits further attention. Along with registration of trials, this closely related strategy has been advocated as a means of avoiding publication bias. It may be possible, however, to go beyond simply planning the logistics of multiple trials and the collection of common data elements to allow pooling of results upon the completion of all trials. It may be possible to go further toward planning the scientific questions to be addressed. As a simple example, by regulation, sex and age (adult versus pediatric) would need to be addressed for a new analgesic. In addition, it would be important to consider indication (emergency department, postoperative, etc.) and dose (cumulative dose, daily dose, need for a loading dose, etc.). How best to design the *series* of studies to address all of these questions, either simultaneously or sequentially, needs further consideration (see Berlin and Colditz, 1999, for a more complete discussion of this issue).

While there are no easy answers to many of these questions, 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 is of the utmost importance and needs 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.

Key Points

- Meta-analysis, if carefully done, is a powerful method that can be used to identify sources of variation among studies and provide an overall measure of effect.

(Continued)

- Combining evidence across diverse study designs and study populations may lead to generalizable results.
- Publication bias, and flaws in the design of component studies, should lead to careful interpretation of meta-analyses.
- Meta-analysis could save considerable time and resources between a research finding and the clinical implementation of a new therapy, by accumulating evidence as it becomes available.

SUGGESTED FURTHER READINGS

- Andrews TC, Reinold 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 5): 236–44.
- Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analysis of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992; **268**: 240–8.
- Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989; **81**: 107–15.
- Berlin J, Santanna J, Schmid CH, Szczech LA, Feldman H. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002; **21**: 371–87.
- Berlin JA, Colditz GA. The role of meta-analysis in the regulatory process for foods, drugs, and devices. *JAMA* 1999; **281**: 830–4.
- Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose–response data. *Epidemiology* 1993; **4**: 218–28.
- Bero LA, Rennie D. The Cochrane Collaboration—preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995; **274**: 1935–38.
- 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.
- Carson JL, Strom BL. The gastrointestinal toxicity of the non-steroidal anti-inflammatory drugs. In: Rainsford KD, Velo GP, eds, *Side-effects of Anti-inflammatory Drugs 3*. Boston, MA: Kluwer, 1992; pp. 1–8.
- Carson JL, Strom BL, Soper KA, West SL, Morse ML. The association of nonsteroidal anti-inflammatory drugs with upper gastrointestinal tract bleeding. *Arch Intern Med* 1987; **147**: 85–8.
- Chalmers TC. Problems induced by meta-analyses. *Stat Med* 1991; **10**: 971–9.
- 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** (suppl 25): 9–26.
- Deeks J, Bradburn M, Localio R, Berlin J. Much ado about nothing: statistical models for meta-analysis with rare events. 6th International Cochrane Colloquium, Baltimore, MD, 1998.
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986; **7**: 177–88.
- Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. *Epidemiol Rev* 1992; **14**: 154–76.
- Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; **316**: 140–4.
- Gabriel SE, Jaakkimainen L, Bombadier C. Risk for serious gastrointestinal complications related to the use of nonsteroidal anti-inflammatory drugs. A meta-analysis. *Ann Intern Med* 1991; **115**: 683–91.
- Hennessy S, Berlin J, 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.
- 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 [see comments]. *BMJ* 1996; **312**: 1563–6.
- Jüni P, Witschi A, Block R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; **282**: 1054–60.
- Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. New York: Van Nostrand Reinhold, 1982.
- Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol* 2002; **55**: 86–94.
- 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.
- Normand S-LT. Tutorial in biostatistics. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* 1999; **18**: 321–59.
- Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. *Control Clin Trials* 2000; **21**: 488–97.
- Sterne JAC, Egger M, Smith GD. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; **323**: 101–5.
- Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-Analysis in Medical Research*. Chichester: John Wiley & Sons, 2000.
- Sutton AJ, Duval S, Tweedie R, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on metaanalyses. *BMJ* 2000; **320**: 1574–7.

- Szczech LA, Berlin JA, Feldman HI. The effect of antilymphocyte induction therapy on renal allograft survival. A meta-analysis of individual patient-level data. Anti-Lymphocyte Antibody Induction Therapy Study Group. *Ann intern Med* 1998; 128: 817–26.
- Szczech LA, Berlin JA, Aradhye S, Grossman RA, Feldman HI. Effect of anti-lymphocyte induction therapy on renal allograft survival: a meta-analysis. *J Am Soc Nephrol* 1997; 8: 1771–7.
- Taragin MI, Carson JL, Strom BL. Gastrointestinal side effects of the nonsteroidal anti-inflammatory drugs. *Dig Dis* 1990; 8: 269–80.
- Webber DM, Montague TH, Bird NP. Meta-analysis of QTc interval- pooling data from heterogeneous trials. *Pharmaceut Stat* 2002; 1: 17–23.

Patient Adherence to Prescribed Drug Dosing Regimens in Ambulatory Pharmacotherapy

The following individuals contributed to editing sections of this chapter:

JOHN URQUHART¹ and BERNARD VRIJENS²

¹AARDEX Ltd, Zug, Switzerland, Biopharmaceutical Sciences, UCSF, San Francisco, USA, and Maastricht University, Maastricht, The Netherlands; ²Pharmionic Systems Ltd, Visé, and University of Liège, Belgium.

INTRODUCTION

The importance of the topic of this chapter arises largely from two factors: (i) prescription drugs are a cornerstone of medical care, based on the continuing flow into the market since World War II of new pharmaceutical products of steadily growing therapeutic power; (ii) it is a basic axiom of pharmacology that all therapeutic drug actions are dose and time dependent.

Point (i) is an historical fact; point (ii) is a basic principle in pharmacology, applicable to all synthetic or natural chemicals that qualify as “drugs,” though the quantitative details differ widely from one drug to another (see also Chapter 4). Point (ii) also applies to the vast majority of drug side effects, except for certain allergic or other reactions that are triggered by a patient’s exposure to even tiny quantities of drug. Most side effects occur in a seemingly dose-dependent manner, and include the most commonly occurring ones (e.g., headache, dizzy spells, nausea, vomiting, constipation, diarrhea, abdominal cramps), which have dose- and time-dependent expressions, just as do therapeutic

effects. Thus, the magnitude of the benefits to be realized from use of a prescribed pharmaceutical depends on whether it is in fact ever taken, on the temporal sequence of doses taken, and on the duration of the patient’s continuation with the drug dosing regimen. After dosing is discontinued – whether on medical advice or on the patient’s own initiative – drug actions fade away, not to return unless at some later time dosing resumes.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

ESSENTIALS OF PATIENT ADHERENCE AND ITS VARIATIONS

The topic of patient adherence is rather simple in overview, but complex in its underlying details. The topic of what patients actually do with prescribed drugs has become a field of research known as *pharmionics*, one of the subdisciplines of the biopharmaceutical sciences.

Pharmionic questions naturally arise whenever patients are prescribed medicines whose procurement and administration are the responsibility of the patient, without direct oversight by health professionals. In contrast, there are certain circumstances, e.g., in-hospital care, when the responsibility for administering medicines is entirely in the hands of health professionals. While health professionals can and do make mistakes (e.g., by failing to administer the prescribed drug in the proper dose, at the proper time, or by administering the wrong drug), the incidence of such errors among health professionals is far lower than among patients. Errors in drug administration made by health professionals are a topic in its own right, for, despite the incidence of professionals' errors being low, their impact is sometimes great, as hospitalized patients are, for the very reasons that cause them to be hospitalized, ill and thus likely to be more vulnerable to errors in their care than are ambulatory patients (see Section IV of this book).

Something of a "middle ground," between patient-administered and professionally-administered medications, is "directly observed therapy," whereby ambulatory patients are asked or, in some instance, required by force of law to come to a clinic to be directly observed taking their scheduled doses. A rising tide of treatment failures and emergent resistance to most of the drugs available for use in tuberculosis treatment provided the motivation for "directly observed therapy." It has had remarkable success in improving cure rates and in greatly reducing the incidence of emergent drug-resistant tubercle bacilli.

PHARMIONICS: OVERVIEW OF DOSING ERRORS MADE BY AMBULATORY PATIENTS

Ambulatory patients frequently make three types of errors in dosing. A patient can, first of all, opt not to *accept* the recommended treatment, and so never procure the medicine from a pharmacist. In some instances they may go so far as to procure the medicine but then take none or only a few doses. Secondly, a patient can commence taking the medicine, but *execute* the prescribed dosing regimen poorly, so that scheduled doses are delayed or omitted, occasionally or frequently. Some dose omissions cluster into multi-day sequences of consecutively omitted doses, called *drug holidays*, potentially creating a transient interruption in drug action. Thirdly, patients can *discontinue* taking the medicine altogether at any time, thereafter taking none or so few doses that they can be categorized as having discontinued execution of the prescribed drug dosing regimen. The high incidence of early discontinuation has given rise to use of the term "persistence," which is defined for each patient as

the time between the first-taken and last-taken doses. It is striking to see the high incidence of short persistence with prescribed medicines that are meant to be taken on a life-long basis, including, for example, antihypertensives and lipid-modifying agents.

Figure 25.1 illustrates these points with data from a large group of patients prescribed one of a variety of antihypertensive drugs. Note the immediate drop in the fraction of patients who are engaged with the dosing regimen, shown by the dashed line. These are the non-acceptors, representing about 4% of the cohort being studied, but often larger in other trial or practice settings. Thereafter, the dashed line gradually declines, indicating decline in the fraction of patients who are still engaged with the dosing regimen. The irregular solid line plots the fraction of patients who have dosed correctly on each consecutive day. For example, if 1000 patients are enrolled in a study with a once-daily dosing regimen, we lose 40 immediately, due to non-acceptance, leaving 960 patients engaged with the dosing regimen, but this number dwindles so that, at day 60 after the start of treatment, it reaches 800 of the original 1000 patients still engaged with the dosing regimen. Of these 800 patients, 700 took the prescribed dose on day 60, thus omitting 100 doses out of 800 prescribed or 12.5% and creating the day-60 gap between the dashed and the solid lines. From day to day the fraction of patients who dose correctly varies slightly, creating the small fluctuations in the solid line. Some patients miss many doses from day to day whereas others miss few or none. As time passes, there is a further dwindling of the number of patients still engaged with the dosing regimen, and the solid line gradually comes closer to the dashed line. This narrowing of the gap between the two lines reflects the fact that patients who omit many doses are more likely to discontinue early than those who dose correctly. Thus those continuing tend, increasingly, to be patients who execute the drug dosing regimen correctly. This process is probabilistic, so there will be some strictly punctual dosers who discontinue early, and some erratic dosers who continue throughout the one-year period of observation, but the proportion shifts toward those who execute the drug dosing regimen correctly as treatment-time lengthens.

TERMINOLOGY

Several key terms are used in the field of pharmionics. "Adherence" is an over-arching term that encompasses all three of the major types of error – nonacceptance, poor execution, and early discontinuation. The term "adherence" provides a certain convenience, but falls short in a critical respect:

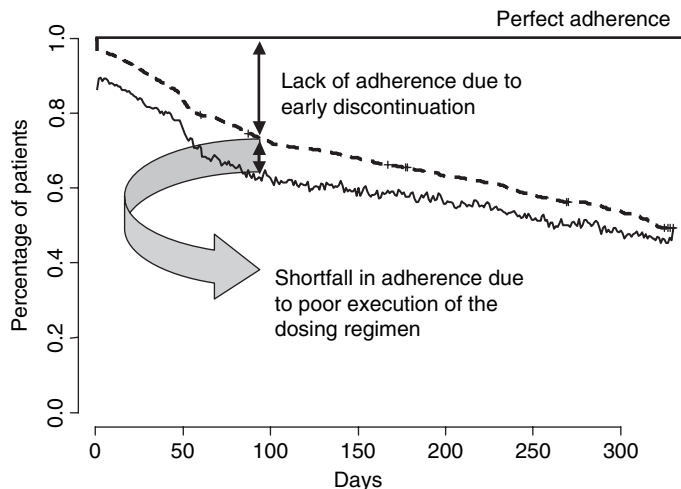


Figure 25.1. Time-course of adherence parameters (acceptance, execution, persistence) from over 4000 patients treated during months to years for arterial hypertension. The various aspects of the figure are discussed in the text. (Copyright © 2006, AARDEX Ltd. Reproduced with permission.)

when someone says that a patient's adherence was poor, you cannot know from that statement alone whether the patient never accepted the treatment plan, whether he/she accepted but executed poorly, or whether (irrespective of the quality of execution) the patient discontinued the medicine sooner than the prescriber intended. Thus, one has to look behind the term *adherence*, to find out in which way(s) the patient fell short in his/her taking of the medicine. "Acceptance," "execution," and "continuation" have become the three key parameters in giving quantitative expression to pharmionic data on what patients do with the medicines they are prescribed.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGIC RESEARCH

A high incidence of the foregoing errors has been more or less evident for many years, but until recently these errors have been seen through the cloudy lens of poor methods that are biased toward estimates that indicated substantially fewer and/or smaller errors than actually occur. This underestimation bias arose because the early methods afforded patients the easy opportunity to censor evidence for delayed or omitted doses. The early methods included clinical judgment, the counting of untaken pills or capsules, questionnaires, interviews, diaries, and measurements of drug concentration in plasma. Intervals between prescription refills are sometimes

used to detect either substantial omissions of prescribed doses or early discontinuation.

Clinical judgment has been aptly dismissed as being "no better than a coin-toss." One reason for the inaccuracy of clinical judgment is that, although the doctor-patient relationship is fundamentally based on trust, many patients are reluctant to disclose the extent to which they have strayed from prescribed dosing regimens. Another reason is the problem of the patient's recall of missed doses and when they occurred.

Counting untaken pills/capsules is easily done, by simply asking patients to bring their medicine container with them to each scheduled visit to the physician, as is typically done in clinical trials. Estimates of adherence based on this method are easily biased upward because patients can and do discard or hoard untaken doses before returning the medicine container to the physician. This behavior can readily be detected, but not corrected for, by giving patients 50–100% more doses than are needed for correct dosing during the interval between visits. When such extra doses are provided, one will find that about 20% of patients return an empty or nearly-empty container. Of course, if the patient returns most or all of the dispensed tablets/capsules, one can reasonably conclude that the patient took few or none of the prescribed doses, but the return of a full or almost full container is an infrequent occurrence. Very strong evidence provided by Pullar *et al.* (1989) was based on a chemical marker's concentrations in blood; this study thoroughly

discredited pill counts as “grossly over-estimating patients” adherence to prescribed drug dosing regimens. Yet the clinical trials community continues to report adherence values based on pill counts.

The results found by questionnaires or interviews are clearly subject to whatever censoring or exaggeration patients wish to apply, over and above the fact that recall of day-to-day details of medicine-taking is almost invariably poor, thus precluding use of these methods to indicate the actual time-patterns of drug intake. Instead, they can usually only convey a qualitative impression of whether a few or many doses were omitted, with little or no information on when dose omissions occurred, and in what sequence between periods of correct or nearly-correct dosing. Of course, if a patient reveals that he/she took few or no doses of prescribed drug, one can reasonably assume that the information is correct, but assertions of correct dosing should be viewed skeptically.

In principle, a diary completed by the patient affords the opportunity to capture timely data that might otherwise be forgotten by the time of the patient’s next visit to the investigator. In practice, however, diary entries tend strongly to be entered long after the fact. A recent study by Stone *et al.* (2002) used a special diary that electronically time-stamped when entries were made, revealing that only 11% of diary entries had a credible temporal relation to the events being recorded.

Measurements of drug concentration in plasma have major interpretative difficulties, notwithstanding their seeming objectivity. The main difficulty arises from the fact that the vast majority of prescription drugs in today’s use have half-lives in plasma of 12 hrs or less. This relatively rapid turnover severely limits the utility of measuring drug concentration in plasma, for two reasons. The first arises from a basic principle in pharmacokinetics: a measured concentration of drug in plasma made at a given moment reflects drug dosing during a prior period of time equivalent to 3–4 times the drug’s plasma half-life, i.e., 36–48 hrs or less for the vast majority of drugs (see Chapter 4). The second is that the 36–48 hrs time interval before the drawing of blood turns out to coincide with a special period in many patients’ dosing histories, wherein, during the 2–3 days prior to a scheduled visit, many patients who ordinarily skip many scheduled doses manage to improve substantially their execution of the dosing regimen. This pre-visit improvement in regimen execution is called *white-coat compliance*, and it imposes a major bias on the interpretation of measured concentrations of drug in plasma as an indicator of the quality of a patient’s usually prevailing execution of the prescribed dosing regimen. Of course, if a measured concentration of drug is found to be zero or

very low, with no evidence for exceptionally rapid clearance of drug from plasma, one can be reasonably certain that the patient took few or no doses of drug during the prior period of time equivalent to 3–4 times the drug’s half-life in plasma. For the relatively few drugs that happen to have a much longer than the usual plasma half-life, a measurement of drug concentration in plasma indicates aggregate drug intake over a correspondingly longer period of time, but still gives no information on when omissions and other dosing errors occurred.

CURRENTLY AVAILABLE SOLUTIONS

ELECTRONIC MEDICATION EVENT MONITORING

A major methodological breakthrough occurred in the 1980s when the technique of *electronic medication event monitoring* (eMEM) was introduced for use in clinical research studies. The essence of the eMEM method is to imbed into the drug package microcircuitry that is connected to one or more micro-switches which detect when the maneuvers occur that are needed to remove a dose of drug from the package. Those maneuvers, which vary from one type of package to another, are referred to as “medication events.” The microcircuitry enters the time of occurrence of each medication event, and stores that information in memory for later transfer to a computer, which analyzes the data on the assumption that each medication event signifies that the prescribed dose was actually taken at the recorded time.

VALIDITY OF THE ASSUMPTION THAT MEDICATION EVENTS SIGNIFY CORRECT DOSING

The validity of this assumption has been tested by using recorded medication events as input to a pharmacokinetic model for each patient whose medication events record has been compiled electronically, and then comparing the resulting projected concentrations of drug in plasma with directly-measured concentrations in blood samples drawn according to a pre-defined sampling plan. The first published study based on use of that maneuver was done during a year-long study of patients being treated with a protease inhibitor for infection with the human immunodeficiency virus. The output of each patient’s pharmacokinetic model was a continuous, year-long projection of the time-course of the concentration of the protease inhibitor in plasma. During the year, blood was sampled at various times, permitting

direct measurements of the plasma concentration of the protease inhibitor to be made at those times. The projected plasma concentrations at those same sampling times were compared with the directly measured concentrations. The differences between projected and directly measured values were very small.

For that degree of correspondence to occur between projected and directly measured concentrations of drug, patients had to have taken the drug in the prescribed amount, at the times recorded by the electronically monitored package. Of course, one can expect, when hundreds of patients have been studied with this technique, to see isolated instances in which eMEM data are biased either upward or downward, depending on mistakes made in the instructions for use of the electronically monitored drug packages, but the available data indicate that such errors are infrequent. The upshot, however, is that the inherently indirect method of eMEM tentatively can be regarded as reliable, pending more data of the type described.

WHAT DO WE LEARN FROM HAVING RELIABLY COMPILED DRUG DOSING HISTORIES?

The foundation of learning from reliably compiled drug dosing histories is that groups of ambulatory patients will generate a wide diversity of temporal patterns of drug exposure. For the most part, these will be dose omissions of various durations, with varying degrees of abruptness in the cessation of dosing and in its subsequent resumption. The clinical correlates of these various temporal patterns of drug exposure create an extensive opportunity to learn about the drug's pharmacodynamics.

COMMON DOSING ERRORS MADE BY AMBULATORY PATIENTS

The most common error is a delayed dose, still taken within the scheduled interval between doses. The next most common error is to omit a single dose. Other dosing errors, in decreasing frequency of occurrence, are: to miss two sequential doses, to miss three sequential doses, and so forth. These, and the patterns described below, are findings from the Pharmionic Knowledge Centre in Visé, Belgium, which contains an archive of data from over 15 000 patients whose dosing histories have been contributed by investigators who have used the AARDEX MEMS[®] Monitors in various clinical studies.

A prominent feature of patients' dosing histories is the higher occurrence of dose omissions with evening doses than

with morning doses, and a higher occurrence of dose omissions on weekends than on weekdays. Another prominent feature is a gradual increase in the frequency of dose omissions as the duration of treatment increases. Still another feature is "white-coat compliance," noted earlier. Obviously, white-coat compliance can result in a clinical picture at the time of a scheduled visit that misleads the physician into concluding that the patient is continuously correctly executing the prescribed drug-dosing regimen. Knowing the full dosing history that precedes a measurement of drug concentration in plasma avoids the misinterpretations prompted by white-coat compliance. Moreover, knowing the full dosing history largely obviates the need for measuring drug concentrations in plasma, except to confirm the pharmacokinetic projections made from dosing history data.

PHARMACODYNAMIC CORRELATES OF COMMON DOSING ERRORS

The pharmacodynamic correlates of these diverse patterns of drug exposure represent a rich, natural experiment in dose ranging that would be difficult or impossible to create and study by purposeful design. Examining the pharmacodynamic correlates of these various patterns of drug exposure thus represents a new chapter in clinical investigation that integrates clinical epidemiology and clinical pharmacology.

The various patterns of dose omissions seen in early trials of a new drug are likely to be those that will occur in later, larger trials, and, beyond that, in routine clinical use of the drug. Thus, there is a simple, pragmatic reason for wanting to know what impact these deviations from the recommended dosing regimen have on the effectiveness and safety of the drug in question. But, first, and foremost in early drug development, the observation of the clinical correlates of these lapses in dosing can help in answering the burning question that inevitably overhangs early drug development, namely ...

IS THE SELECTED DOSING REGIMEN OPTIMAL?

"Optimal" is more a figure of speech than a term carrying a precise definition. The basic reason is that the pharmionically-informed selection of a recommended dosing regimen represents a compromise between desirable and undesirable consequences both of low-end and high-end dosing regimens. In the simplest case, if patients exposed to one-quarter to one-half of a protocol-specified level of drug exposure are still showing essentially fully developed effects of the drug, one might reasonably suspect that the

protocol-specified dosing regimen has been set considerably higher than necessary. Such a finding should not, for reasons discussed below (under the heading “Limits on what one can learn from the clinical correlates of variable dosing”), be considered as definitive evidence, but rather as a strong “red flag” that the protocol-specified dose has been set to a level higher than necessary (see Case Example 25.1).

CASE EXAMPLE 25.1: HOW MUCH ADHERENCE IS ENOUGH?

Background

- Doxycycline hyclate 100 mg, given orally twice daily for 7 days, is the generally accepted standard of care for chlamydial infections of the male urethra or the lower genital tract of females.

Question

- How much does adherence alter responsiveness to doxycycline?

Approach

- A study carried out by the Public Health Department of the State of Alabama (USA) sought to examine the impact of poor compliance with the prescribed dosing regimen for doxycycline (100 mg, given orally twice daily for 7 days) on outcomes of chlamydia treatment.
- Prescribed doxycycline was supplied to the trial participants in electronically monitored drug packages.

Results

- Patients who took as few as 25% of the prescribed doses of doxycycline appeared to have had successful outcomes of treatment in no significantly lesser percentage than those who took all or virtually all of the prescribed doses.
- The outcomes of treatment appeared to be independent of adherence!
- This is *prima facie* evidence that the prescribed dose could be substantially reduced.

Strength

- The patients performed the natural experiment of underdosing and demonstrated that as little as 25% of the prescribed dose appeared to be effective in treating

chlamydial infections of the male urethra and female lower genital tract.

Limitations

- It is a natural experiment, and so should be viewed as a “red flag” rather than as definitive proof about how the dosing regimen could be revised.
- The study came late in the commercial history of doxycycline, after its key patent protection had expired and its pricing came down to the usual multi-source (“generic”) level. Thus the economic advantages of cutting the dose at this point in the product’s history are very small, but possibly still important in some locations where health care budgets are severely constrained.
- It is not clear whether reducing the dose by a factor of 2–3 would reduce side-effect problems, e.g., photosensitivity.

Summary Points

- The “natural experiment” can indicate the possibility of reducing recommended doses.
- It is preferable, from the consumer’s perspective, to make this discovery early rather than late in a pharmaceutical product’s commercial lifetime.
- Pharmaceutical product developers can avail themselves of the natural experiment early in drug development, before the recommended dosing regimen has been set, and before associated pricing has been set. Doing so can avoid the adverse economic consequences of a postmarketing, post-pricing reduction in actually used doses.

CURRENT METHODS OF SELECTING RECOMMENDED DOSING REGIMENS FREQUENTLY LEAD TO MAJOR ERRORS

The question of dose-optimization has taken a strong turn in the past 25 years that focuses much more attention than ever before on the question of whether a recommended dosing regimen can be reduced. There are four basic reasons: (i) premarketing studies are inevitably limited in size and duration (see Chapter 1), the dose evaluated is to some degree an educated guess, and it is quite possible for a sponsor to choose the wrong dose premarketing for final evaluation and testing; (ii) drug prices have escalated markedly during the past 25 years, creating correspondingly increased incentive for purchasers of the product to find ways to reduce the dose or the dosing-frequency, which, given that most

pharmaceutical products are priced on the basis of drug quantity, correspondingly cuts the cost of treatment; (iii) the coalescence of many medical practices into managed care organizations (MCOs), which use formulary committees to review each new pharmaceutical product added to the organization's formulary, exposes each new pharmaceutical and its associated dosing regimen to a high level of scrutiny; and (iv) many MCOs have the ability to create and enforce prescribing and dispensing guidelines, which can include the use of a lower or less frequently administered dose than recommended, based on the formulary committee's review of the available evidence.

The net result of these various forces, and perhaps others, has been that 22% of new drugs that entered the market between 1980 and 2000 underwent at least a 50% reduction in either their label-recommended dose or their most commonly prescribed dose. Moreover, the incidence of such postmarketing dose-reductions rose during the decade after 1990. Postmarketing dose reductions occur after the product's pricing has been set, which effectively makes the product an economic hostage to the stability of its recommended dosing regimen, given that a 50% cut in the actually-used dose means an approximately 50% reduction in revenue that the manufacturer derives from the product.

The upshot of the foregoing is that there is a degree of uncertainty about whether the recommended dosing regimen for each new pharmaceutical is close to optimal. Obviously, if the recommended dosing regimen has been set too high, it may nullify clinical consequences of single or multiple dose omissions.

BASIC CONSIDERATIONS IN DEFINING THE OPTIMAL REGIMEN

An optimal dosing regimen provides enough "forgiveness" so that drug action is not seriously interrupted by the more commonly occurring dosing errors – missing a single dose, or two sequential doses. It is also a dosing regimen that does not provide so much forgiveness that would give would-be dose-cutters a convenient handle to lower the dose or dosing frequency, compared to what has been recommended by the drug developer.

The definition of "forgiveness" is: the duration of therapeutically effective drug action after a last-taken dose *minus* the recommended interval between doses. Fortunately, there are enough examples of measured forgiveness to indicate how widely this parameter can vary from one product to another.

For example, the forgiveness of the low-dose, combined estrogen/progestin oral contraceptive (the most widely used

form of the oral contraceptive "pill") is only 12–24 hrs, depending on whether the calculation is based on the UK or the US labeling of these widely used products. So by 36–48 hrs after a last-taken oral contraceptive pill the risk of breakthrough ovulation and conception begins to rise. At the opposite end of the forgiveness spectrum we find bendrofluthiazide, which is the antihypertensive thiazide diuretic that is most widely used in the UK market, with a forgiveness in excess of 5 days, based on its once-daily dosing regimen and its measured duration of 6.3 days of antihypertensive action after a last-taken dose.

Just considering these two products, one can see that a delay in taking the once-daily oral contraceptive pill from morning to bedtime creates an extended interval between doses of about 40 hrs, which may open a time-window within which the risk of breakthrough ovulation can rise, putting the patient in jeopardy for conception. Not surprisingly the conception rates that one measures in clinical trials are very closely linked to the punctuality with which the once-daily dosing regimen is executed, giving rise to the data in Table 25.1, which come from a recent review of the contraceptive field by the US Centers for Diseases Control.

Thus, there is a 50-fold degradation in contraceptive protection between "perfect" use and "typical" use. This huge difference is emblematic of a too-unforgiving product, which, in effect, requires a degree of consistent punctuality that only a relatively small minority of patients maintain. In contrast, a 5-year steroidal implant was able to achieve the lowest conception rate on record for any of the steroidal contraceptives. This remarkable degree of effectiveness is attributable, with reservations discussed below, to the continuity of steroidal action maintained by the continuous, controlled release of its contraceptive steroid, norgestrel.

BEWARE OF AVERAGES

Note that in the foregoing discussion we have used an average, or mean, value for the drug's duration of therapeutically effective action after a last-taken dose. A major problem with averages is that half of the patients whose data contribute to the average value had lesser values than

Table 25.1. Annualized conception rate

Mode	Perfect use	Typical use
Daily pill	0.1%	5.0%
5-year implant	0.05%	0.05%

Data from the US Centers for Disease Control (1999).

the average, while the other half had greater values than the average. Of course, if the data are skewed to one side of the average value, then the view is more complex, so let us stick in what follows to a hypothetical example of a strictly symmetrical (Gaussian) distribution around the average.

For example, if the drug regimen is to call for once-daily dosing, then in the ideal case we need to assure that every patient who doses correctly, at precisely 24-hr intervals, will achieve and maintain continuity of drug action. It does not suffice to recommend once-daily dosing if the average post-dose duration of drug action is 24 hrs, because half of the patients will have smaller values and thus will experience gaps in drug action when doses are taken at 24-hr intervals. We therefore have to examine the range over which the measured post-dose duration of drug action varies. The conventional parameter that captures this range is the standard deviation of the mean. If the mean minus one standard deviation equals 24 hrs, then one-sixth of patients treated with a once-daily dosing regimen will be unable to achieve continuity of drug action with punctual dosing. If the mean minus two standard deviations equals 24 hrs, then only 2.5% of patients will be unable to achieve continuity of drug action with punctual dosing. If we wanted to guarantee that every single patient achieves continuity of drug action with punctual dosing, then we would have to borrow from the principles of highest-precision manufacturing methods and use the mean minus six standard deviations.

So the question then becomes: what is the magnitude of the standard deviation of the post-dose duration of drug action? The usual answer with physiological parameters, such as post-dose duration of drug action, is about 30% of the average. Clearly, it is impossible with such a large standard deviation to consider the mean minus six standard deviations, as that will be a negative number. More realistically, the mean minus two standard deviations could be considered acceptable. Thus, for once-daily dosing one would need to have an average post-dose duration of action of at least 60 hrs, with a standard deviation of 18 hrs. Thus, two standard deviations would be 36 hrs, and the average minus two standard deviations would be $60 - 36 = 24$ hrs.

TWO EXAMPLES

If we apply this reasoning to the low-dose estrogen/progestin oral contraceptive, what might we expect to find? Let us use the 48-hr post-dose duration action that the FDA has implicitly put into the product labeling as the basis for informing patients what to do when they have missed a pill. Let us use the estimate that the standard deviation will

be 30% of the average, i.e., 14.4 hrs. Thus the average minus two standard deviations will be $48 - 28.8 = 19.2$ hrs. Thus, a small fraction of patients, despite punctual dosing, could be expected to have “negative” forgiveness: $19.2 - 24 = -4.8$ hrs, and thus be vulnerable to breakthrough ovulation and conception (see Case Example 25.2).

CASE EXAMPLE 25.2: HOW MUCH ADHERENCE IS ENOUGH IN ANOTHER FIELD OF THERAPY?

Background

- Combined estrogen/progestin oral contraceptives underwent a major dose reduction in the early 1970s, in response to evidence that the high-dose products carried an unacceptable risk of arterial thrombotic events.
- Although the low-dose forms have largely, if not entirely, avoided these hazards they carry a much greater risk of unwanted conception than did the high-dose products.
- The high-dose products could maintain steroidal blockade of ovulation in the face of several days' omission of doses, whereas the low-dose products could not.
- A series of studies, carried out during the 1980s, showed that the risk of “breakthrough” ovulation began to increase within 36–48 hrs after a last-taken dose of the low-dose “pill.” When the US Centers for Diseases Control reviewed the field of family planning in 1999, their review included the data presented in Table 25.1.

Questions

- How does one define how much adherence is enough for a patient-administered pharmaceutical product?
- How does one represent that information in product labeling?

Approach

- Studies run during the 1980s were based on the assumption that a controlled substitution of one or more placebo pills for active pills was the equivalent of a patient's spontaneous omission of one or more sequential, once-daily doses of the oral contraceptive.
- For ethical reasons, the studies were run in volunteers who had previously had a tubal ligation, so that they were unable to conceive, but still ovulated.

- The endocrinological “signature” of imminent ovulation is a sudden, sharp rise in plasma levels of the pituitary hormone LH (luteinizing hormone).
- Five such studies were run in different centers in Western Europe and North America, with minor differences in study design and methods.

Results

- The ovulating surge of LH began to occur in a few volunteers as early as 36 hrs after a last-taken dose of the combined, low-dose, estrogen/progestin oral contraceptive. By 48 hrs since the last-taken dose, a substantial fraction of volunteers showed the ovulating surge of LH.
- The data are complicated by the fact that the risk of ovulation varied throughout the 28-day period between successive menses.
- The key question then became: what information should patients be given in product labeling about what they should do when it became apparent that they were late in taking the next-scheduled dose, or that they had missed one dose completely, or that they had missed two or more sequential doses?
- The use of barrier contraceptives – diaphragm, foam, condoms – was deemed to be a key counteractive maneuver, with patients instructed to use 7 days’ barrier contraception following the recognition that a dose omission had occurred.
- The exact steps to take depended on whether the dose of the pill was delayed but not omitted, omitted on only one day, omitted on 2 days, or omitted on more than two days.

Strengths

- The measurement of the time after a last-taken dose until contraceptive action stops allowed for a rigorous quantification of drug effect in the setting of poor adherence.
- The recognition that a dose has been delayed or omitted is facilitated by the calendar-pack mode of packaging used for oral contraceptives. (Attempts to put “beepers” into oral contraceptive or other pharmaceutical packages have proven to be a widely unacceptable maneuver for many reasons, notwithstanding its intuitive appeal.)

Limitations

- The details are complicated by a number of factors: (i) different lengths of dose omission; (ii) variations in

the likelihood of breakthrough ovulation during the 28-day period between sequential menses; (iii) uncertainties about the effectiveness of barrier contraceptives; and (iv) uncertainties about the effectiveness of each of the recommended maneuvers.

- The British and American regulators differed on how much time after a last-taken dose they should allow patients before recommending that they institute barrier contraception. The British said 36 hrs, the Americans said 48 hrs. No one knows which of these recommendations is the better.

Summary Points

- The low-dose, combined estrogen/progestin oral contraceptives are quite unforgiving products, requiring strict dosing for full contraceptive effectiveness, as shown in Table 25.1.
- Extensive research has provided a reasonably sound basis for labeling information for patients on what to do when they find that they have delayed or missed doses.
- No one knows the effectiveness of the information provided on “what to do when you miss a dose.”

If we examine the situation with a highly forgiving agent, e.g., bendrofluthiazide, we might expect that, on the same assumptions previously used, the standard deviation of 30% of the mean post-dose duration of action of 6 days would be 1.8 days. The value of the 6-day mean minus two standard deviations would be $6 - 3.6 = 2.4$ days, which signifies 1.4 days (2.4 days minus 1 day) of forgiveness (33.6 hrs), indicating that most patients, but not all, could miss two consecutive daily doses and still maintain continuity of drug action.

In general, the crucial problem in setting a recommended dosing regimen is to strike the right balance between two conflicting requirements. On the one hand, the product should provide enough forgiveness to allow the vast majority of patients who use the product correctly to have it provide continuity of action, i.e., the mean forgiveness minus two or more standard deviations. On the other hand, the product should have a convenient dosing regimen, i.e., once- or twice-daily, and not provide so much forgiveness that it tempts would-be dose-cutters to attack the recommended dosing regimen.

Here it is necessary to recognize that once-daily dosing has been greatly oversold as a “solution to the compliance problem.”

ADHERENCE, FORGIVENESS, AND CONTINUITY OF DRUG ACTION WITH ONCE- AND TWICE-DAILY DOSING

If the measure of adherence is considered to be the percentage of prescribed doses taken, then once-daily dosing almost invariably shows a slightly higher percentage of prescribed doses than does twice-daily dosing. The therapeutically more relevant question, however, is not how many doses are or are not taken, but what is the comparative impact of missed doses in the two regimens on the continuity of drug action? A key point in this regard is that the pharmacokinetic equivalent of a single missed once-daily dose is between two and three sequentially omitted twice-daily doses. Thus, an important parameter is the probability of two or three twice-daily doses being *sequentially* omitted, versus the probability of missing a single once-daily dose. Indeed the probability of sequential omission of 2–3 twice-daily doses is less than half of the probability of omission of a single once-daily dose. This difference in probability of pharmacokinetically equivalent errors in dosing is one of the factors that makes it likely that a twice-daily regimen is superior to a once-daily regimen.

Of course, the more important factor is maintenance of therapeutic levels of drug action, not the concentration of drug in plasma. The post-dose duration of therapeutic drug action is, as we have seen, drug-specific and dependent on the pharmacometric properties of the formulated drug. These have to be measured experimentally: they cannot be predicted from pharmacokinetic data alone. The upshot is that judging the comparative superiority of one dosing regimen over another requires knowledge of the formulated drug's post-dose duration action, plus knowledge of the comparative probabilities of the various patterns of dose omission.

LIMITS ON WHAT ONE CAN LEARN FROM THE CLINICAL CORRELATES OF VARIABLE DOSING

Drug dosing histories arise from a variety of behavioral factors that lead individual patients to engage in a wide range of variable dosing, mostly expressed as variable intervals between doses. It is noteworthy that clinical pharmacologists test drugs usually by varying the dose but holding constant the interval between doses, whereas patients usually hold the dose constant and vary the interval between doses. These are probably not equivalent maneuvers for many drugs, so that one might look at the clinical correlates of the “natural experiments” in dose-ranging being run by patients as potentially providing complementary information to the purposefully designed, randomized,

controlled experiments in dose-ranging being run by clinical pharmacologists. In view of the rising frequency of over-estimated dosing requirements, the addition of the “natural experiments” to preclinical drug development could play a key role in facilitating the early exposure of dosing regimens that call for too high or too-frequent doses.

A key advantage of the “natural experiment” is that it is happening whenever patients have the responsibility for their own dosing, needing only the compilation of the patients' drug dosing histories and their concomitant drug effects. Moreover, knowing these dosing histories and their concomitant pharmacodynamics reflects what many patients actually do with prescription drugs, and so has the pragmatic value of showing the consequences of common patterns of variable dosing that can be expected to occur in routine use of the resulting product.

A disadvantage of the “natural experiment,” however, is that the dosing patterns that occur do not arise by purposeful design, with randomization and strict placebo controls, and so are categorized as “observational” phenomena. As such, one does not have the security that randomized, controlled trials provide in the inference of causality. Thus, the clinical correlates of various spontaneously occurring patterns of drug exposure may or may not represent a causal link between the dosing pattern in question and its associated pharmacodynamics.

For example, in the case of the strikingly low conception rate found with the 5-year steroidal implant, the results might have been due, at least in part, to the fact that the implant was a product associated with rather heavy mid-cycle bleeding, which may have substantially discouraged copulation or, when copulation occurred, the ascent of ejaculated spermatozoa, without which conception will not occur.

Several factors, however, help to clarify what is happening. First, other means of providing continuity of exposure to other types of contraceptive steroids have approximately the same high degree of contraceptive effectiveness and superiority over typically used oral contraceptive pills, thus strengthening the conclusion that continuity of exposure to contraceptive steroids is the causative factor in achieving an ultra-low conception rate. Second, the more commonly occurring deviations from the prescribed dosing regimen also recur, so that, in principle, one has the opportunity to observe repetition (or re-challenge as it is sometimes called) of the clinical events that accompany a particular dosing pattern. Third, causality flows unidirectionally in time: if A causes B, then A must precede B. Thus, one should consistently observe that a particular dosing pattern uniformly precedes the clinical events in question. Fourth, within

ethical limits, there are opportunities to repeat the natural experiment under suitably controlled and randomized conditions, using certain of the more commonly occurring dosing patterns that appear to be clinically important, so as to have the protection in the inference of causality that is provided by randomization and blinding. Fifth, a current area of biostatistical research and methodological development, called “causal inference,” focuses on developing ways and means to strengthen the inference of causality from observational data, such as those we are considering here. Thus, there are multiple ways to confirm or reject causal interpretation of the pharmacodynamic correlates of the “natural experiment” in dose-ranging provided by patients’ variable execution of prescribed drug dosing regimens.

There is a sixth factor to consider, as well: patients are, on their own authority, constantly creating many natural variations in drug exposure that would be categorically unethical to impose by purposeful design. The ethical issues in repeating the natural experiment in properly controlled circumstances depend on the clinical situation, drug, and disease. For example, it would clearly be unethical to impose purposefully a lapse in a tuberculosis-infected patient’s exposure to an anti-tubercular drug of already proven effectiveness. In contrast, one probably could justify the imposition of a purposeful lapse in dosing with an anti-hypertensive agent used to treat mild hypertension, or a lipid-lowering agent used to make beneficial changes in the various fractions of cholesterol, or an osteoclast-inhibiting agent in the prevention of osteoporosis. One factor that may help to alleviate ethical concerns is evidence that the dosing pattern in question be one that the patient in question recurrently executes on his/her own authority. One could make the case that it would be ethically justifiable to study the consequences of such a dosing error because the results would be sufficiently and uniquely beneficial to that patient’s subsequent clinical management to justify the experiment. These, however, are judgments with which ethical committees must wrestle, with each having its own threshold for judging ethical acceptability against scientific imperatives.

WHERE WE STAND TODAY

Table 25.2 lists the different kinds of experimental designs that have been used to provide information on the clinical consequences of a given dosing pattern. The listing is, more or less, in descending order of confidence in the inference of causality between a given pattern of dose omissions and the associated clinical events.

Is Table 25.2 a large or a small list of drugs whose actions have some evidence for their dependence on patients’ variable execution of prescribed drug dosing regimens? There are approximately 500 drugs in frequent clinical use, each with its own relations between dose and concentration in plasma (which is included in the purview of pharmacokinetics) and between concentration in plasma and the magnitude and time-course of the drug’s actions (which is included in the purview of pharmacodynamics). In that sense, Table 25.2 is small. On the other hand, the topic of how patients’ variable execution of prescribed drug dosing regimens impacts on drug actions has only been under consideration since about 1970, and reliable methods for compiling drug dosing histories in ambulatory patients have only been available since about 1990. Against that background, it is reasonable to conclude that Table 25.2 represents both a reasonably good progress report and a template for future studies with other drugs.

SOME GENERALIZATIONS ABOUT THE IMPACT OF VARIABLE ADHERENCE TO PRESCRIBED DRUG REGIMENS

One generalization is clear: drugs do not work in patients who do not take them. That simple fact means that drug actions will be nil in patients who did not accept the principle of treatment, and so never started taking the drug. Many drugs have been tested in randomized, placebo-controlled trials, so that relatively good information exists for most drugs on what happens to patients in the placebo arm of such trials. One can reasonably expect non-acceptors to have a clinical course that approximates what was found in patients randomized to the placebo arm of such trials. It is, however, an open question whether non-acceptance and treatment with a placebo have indistinguishably different effects, for it may be that non-acceptors represent a strongly biased sample of patients enrolled in a study, e.g., toward the low end of disease severity. Non-acceptors have yet to be studied carefully, so the best one can say at present is that drug effects will not occur in non-acceptors.

In the case of early discontinuers the story is somewhat more complicated because each such patient has had a period of variable duration during which he/she took drug, more or less in conformity to the prescribed dosing regimen, before stopping dosing. The magnitude of drug action during the period in which the patient took the drug will, in general, depend on the quality of execution of the prescribed dosing regimen – larger for those who closely followed the regimen, smaller for those who omitted many doses. But then, after dosing has been discontinued, there will be a period of

Table 25.2. Drugs whose actions are or appear to be compliance-dependent

<i>Placebo-substitution-for-active-drug (PSA) study results</i>
Combined estrogen/progestin oral contraceptive steroids
Atenolol
Betaxolol
Trandolapril
Enalapril
Amlodipine
Diltiazem
Bendroflumethazide
Nifedipine
Paroxetine
Sertraline
Fluoxetine
<i>Comparison with assured exposure</i>
Penicillin
Sulfadiazine
Norgestrel
Anti-tuberculosis drugs
<i>Successful measurement-guided intervention in failed therapy</i>
Triple-drug therapy for arterial hypertension
<i>Drug regimen intensification studies</i>
Insulin
Risedronate
<i>Compliance-stratified outcomes of randomized, placebo-controlled drug trials</i>
Cholestyramine
Gemfibrozil
<i>Observed clinical correlates of variable compliance</i>
Warfarin
Estradiol
Phenytoin
Carbamazepine
Pilocarpine
Cyclosporin A
Azathioprine
Allopurinol
Inhaled corticosteroids for asthma
Anti-retroviral drugs

uncertain length during which drug actions gradually fade away, finally bringing the patient more or less to the situation of the placebo recipients in the drug's confirmatory trials. How long that takes is another drug-specific story that must either be determined by experiment or assumed on the basis of available pharmacodynamic data.

The most complicated part of the story are the consequences of the various patterns of dose omissions, especially the drug holidays, which entail three or more days of consecutively omitted doses. Reasonable predictions of what may occur can be based on results of research studies in which measurements have been made of the time, after a last-taken dose, required for drug actions to fall to negligible levels, and how long it takes, when dosing resumes, for

drug actions to rise again to therapeutically useful levels. Few drug researchers have integrated the dosing history findings, published and reviewed in recent years, into their experimental planning, so there is still a general paucity of data from which to predict the clinical and economic consequences of one or another commonly occurring deviation from the prescribed dosing regimen. An additional complicating factor is the lurking possibility that the recommended dosing regimen is substantially suboptimal, as discussed earlier.

While each situation has to be understood in terms of that drug's pharmacometrics (i.e., its pharmacokinetics, pharmacodynamics, and pharmacokinetics), the message is clear that: (i) recommended dosing regimens are not sacrosanct;

(ii) there are strongly negative economic consequences of getting the dose set too high; and (iii) overdosing potentially carries a certain risk of incurring more, or more severe, side effects than would be the case with a more modest dose. Most drugs that qualify as pharmaceuticals offer a wide-enough safety margin between therapeutic and toxic doses that a factor of 2–4 in dose overestimation is, in most instances, more of an abstract concern than a concrete one.

REBOUND EFFECTS AS DRUG HOLIDAYS BEGIN

One of the potentially adverse consequences of drug holidays is their triggering of hazardous rebound effects. The best known and most extensively studied of these concerns the time-course of events after discontinuation of β -adrenergic blocking agents of the class that lack intrinsic sympathomimetic action (ISA) – so-called non-ISA beta-blockers. Such drugs are in wide use, and they include atenolol, propranolol, metoprolol, and others, widely used in the treatment of coronary artery disease, hypertension, and congestive heart failure. Chronic use of these drugs leads to upregulation of beta-receptors. The abrupt cessation of dosing of these agents thus leads to rebound effects, which involve exaggerated sympathetic nervous effects that increase the risk of coronary occlusion or vasospasm, and heightened aggregation of platelets, leading to myocardial infarction or the onset of angina pectoris. The prevention of such events is, of course, the rationale for prescribing non-ISA beta-blockers, and their preventive ability has been confirmed in various ways in many randomized, controlled trials. Of course, in the midst of such trials, some patients were doubtless engaging in drug holidays of varying durations between 3 and 5 days. So, in the wake of the onset of a drug holiday, the holiday-taker is transiently at elevated risk of myocardial infarction or the onset of angina pectoris, though the magnitude of that risk is uncertain. What we know with reasonable certainty is that about 50% of patients will have at least one drug holiday per year, with smaller percentages having multiple holidays, such that one can expect an average of two holidays per patient per year. These can be expected to add to the overall tally of coronary events in the treatment arm of the trial, but not in the placebo arm.

The importance of drug holidays is that potentially holiday-triggered rebound effects probably create coronary events in treated patients that diminish the estimation of the coronary-prevention actions of correctly taken non-ISA beta-blockers. This view, of course, is a qualitative one, not a quantitative one.

How would one study this phenomenon so as to put it on a firmly quantitative basis? One way is to look experimentally at the consequences of abruptly stopping the dosing of non-ISA beta-blockers. Such studies were, indeed, done in the relatively early days of beta-blocker use. At least one such study had to be terminated early because of clearly adverse consequences of rebound effects. A more subtle way to stop the patient's exposure to non-ISA beta-blockers is to do a blinded, randomized substitution of placebo for active drug. In such studies, placebos and active drug look alike to the patient and investigator, but at some moment in time, determined by the protocol and to which the investigators are blinded, the switch is made. But in view of all that we know about the adverse effects of suddenly stopping a non-ISA beta-blocker, most if not all ethical committees would not permit such a study to be done. That being so, the alternatives become doing such a study in animals thought to have sympathetic nervous responses similar to those of humans, or relying on the correlates of the "natural experiment" by collecting dosing histories in ambulatory patients who have been prescribed non-ISA beta-blockers for valid medical reasons, and trying to capture data on the sequence of drug holidays followed by adverse cardiovascular events.

This kind of study, which nicely represents the integration of clinical epidemiology and clinical pharmacology, is difficult or impossible to do unless physiological data collection can be organized so as to capture reasonably comprehensive data on an event that will occur sporadically in most patients, although there are some patients who have several drug holidays per month.

Such a study is much easier in patients who have implanted pacemakers or cardiac defibrillators of the most recently introduced designs, for they collect continuous electrocardiographic data that can indicate myocardial ischemia, signs of impending cardiac arrhythmias, or actual cardiac arrhythmias. This kind of electronic capture of comprehensive data on the effects of drug exposure and the consequences of certain patterns of variable drug exposure gives a glimpse of the future when presumably a greater diversity of electronically captured pathophysiological data will be possible.

Pharmacoepidemiologists have been slow to realize that the cessation and resumption of drug actions, as patients go into and emerge from drug holidays, are potential sources of adverse events – not only from rebound effects, but also from abrupt resumption of dosing of drugs that require gradual increases in dose when treatment is started (so-called "first-dose effects"). These considerations, in turn, reveal how little we know about the consequences of abruptly stopping and abruptly resuming the dosing of drugs. Such

paucity of knowledge arises from a peculiarly one-sided view that clinical pharmacologists have maintained in drug studies, namely to study intensely the onset of drug action in drug-naïve patients but to ignore and not even collect data on the responses to sudden cessation of drug dosing, not to mention subsequent abrupt resumption of dosing.

THE FUTURE

In ambulatory pharmacotherapy, prevalent underdosing, in various temporal patterns, creates a challenging and diverse series of problems. Perforce, it is a major aspect of ambulatory therapeutics and drug trials, notwithstanding its having been long neglected, and still ignored by many. The general progression in pharmaceutical innovation has been, and inevitably will continue to be, toward steadily more powerful agents, thus heightening the importance of having sound pharmionic information and acting upon it in labeling, dosage form and drug regimen design, and clinical practice.

CASE EXAMPLE 25.3: DISTINGUISHING PHARMACOLOGICAL NON-RESPONDERS FROM NON-ADHERERS

Background

- Drugs do not work in patients who do not take them.
- The fairly frequent occurrence of drug non-responsiveness represents an unknown mix of non-adherence and what one might call genuine pharmacological unresponsiveness.
- In the field of hypertension, the syndrome of “drug unresponsive hypertension” is a recognized entity, usually defined as the patient’s failure to respond to so-called “triple therapy,” i.e., three-drug treatment.
- Failure to respond to triple therapy often triggers a clinical investigation into the possible causes of drug unresponsive hypertension.

Question

- Among patients who have failed to respond to triple antihypertensive treatment, what proportion are clinically unrecognized non-adherers?

Approach

- The institution of electronic monitoring of patients’ dosing can trigger, once the patients recognize

that their dosing is under quantitative scrutiny, the commencement of correct dosing, so that the sudden beginning of blood pressure control can be interpreted as evidence that the patients had not previously been taking the medicines.

- A leading Swiss hypertension clinic, to which patients are referred from a large catchment area for diagnostic evaluation, began to precede the usual diagnostic workup of “drug refractory hypertension” with 60 days of electronic compilation of drug dosing histories in patients who had failed to respond to triple therapy.

Results

- Approximately half of the first group of 50 patients so studied turned out to be non-adherers rather than pharmacological non-responders.
- Some of them were documented not to be taking any of the prescribed medicines. Others commenced dosing once it became clear that their dosing would be under close scrutiny. Some of the latter developed postural hypotension after they commenced taking all three prescribed drugs, as their hypertension did not require such strong pharmacological intervention.

Strength

- The first such study in effect cast the reliable measurement of adherence in a new light, i.e., that it can be a diagnostic maneuver that distinguishes pharmacological non-responders from non-adherers.

Limitations

- Non-adherence was not documented prior to initiating the study.
- The investigators interpreted the onset of drug responsiveness following the initial interview and discussion as *prima facie* evidence that the patient’s prior unresponsiveness to prescribed antihypertensive drug treatment had been due to clinically unrecognized non-adherence.

Summary Points

- Roughly half of the cases of drug unresponsive hypertension, on the basis of limited work done to date, may be due to clinically unrecognized non-adherence.
- It seems reasonable to conclude that clinically unrecognized non-adherence will prove to be responsible for a substantive fraction of drug non-responsiveness in other therapeutic fields.

- When drug response is used as a diagnostic maneuver, adherence should be reliably measured to avoid the misinterpretation of clinically unrecognized non-adherence as pharmacological non-response.
- Using patients who have failed to respond to a prior course of drug therapy as subjects for the trial of a new drug is made problematic by the likelihood that a relatively high fraction of non-responders to the first treatment are non-adherers who will carry their non-adherence over to the trial of the second agent.

As previously discussed, the reader may regard Table 25.2 as being a long list of drugs that presently have some form of evidential basis for concluding adherence-dependent drug actions or outcomes of prescribed ambulatory drug treatment. From another perspective, it is a remarkably short list, considering the size of the pharmacopoeia and the basic pharmacological fact that all drugs have dose- and time-dependent actions: in effect, essentially all drugs will sooner or later find places in Table 25.2. Many of the medicines based on these drugs are used for minor indications, as “comfort medicines,” in which case variable adherence is probably clinically unimportant, but adherence and persistence with many other medicines are medically important, and sometimes crucial (see Case Example 25.3). Sound therapeutic analysis can separate the medically important drugs, and sound pharmionic and pharmacometric analyses can sort out which drugs pose substantive hazards when underdosed in certain patterns.

Recognition of the magnitude and diversity of patients’ misuse of prescription drugs has thus led to new methods, new concepts, new terminology, and a new path to the ever-present goals of rational pharmacotherapy and full-disclosure labeling of pharmaceuticals used in ambulatory care. Its application will hopefully continue to expand in the future.

Key Points

- The topic of patient adherence is rather simple in overview, but complex in its underlying details.
- On their own initiative, ambulatory patients engage in a wide variety of patterns of use and misuse of the medicines they are prescribed, usually referred to as “non-adherence.”

- The most important type of non-adherence is early discontinuation of dosing, otherwise known as “short persistence with the prescribed dosing regimen.”
- Early discontinuation of medications:
 - halts drug action, and thus whatever beneficial effects the drug may have induced up to the point of discontinuation,
 - halts the manufacturer’s revenues from the sale of the product to the patient, and
 - wastes whatever costs were incurred in tests and other maneuvers to prepare the patient to start taking the drug in question.
- Until the advent of *electronic medication event monitoring*, it was impossible to know what ambulatory patients actually did with their prescribed medicines.
- Electronic monitoring data have created a new field of research within the biopharmaceutical sciences, called *pharmionics* – the study of what patients do with prescription drugs.
- Pharmionics integrates clinical epidemiology and clinical pharmacology for better understanding of the effects of what patients do with prescriptions and to improve the outcomes of treatments with medications.
- The cessation and resumption of drug actions, as patients go into and emerge from drug holidays, are potential sources of adverse events.

SUGGESTED FURTHER READINGS

- Benet LZ, Øie S, Schwartz J. Design and optimization of dosage regimens; pharmacokinetic data. In: Hardman JG, Limbird LE, Molinoff PB, Ruddon RW, Gilman AG, eds. *Goodman & Gilman’s The Pharmacological Basis of Therapeutics*, 9th edn. New York: McGraw-Hill; 1996; pp. 1707–92.
- Cramer JA, Mattson RH, Prevey ML, Scheyer RD, Ouellette VL. How often is medication taken as prescribed? A novel assessment technique. *JAMA* 1989; **261**: 3273–7.
- 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.
- Feinstein AR. On white-coat effects and the electronic monitoring of compliance. *Arch Intern Med* 1990; **150**: 1377–8.
- Frantz S. Playing dirty. *Nature* 2005; **437**: 942–3.

- Gilligan DM, Chan WL, Stewart R, Oakley CM. Adrenergic hypersensitivity after beta-blocker withdrawal in hypertrophic cardiomyopathy. *Am J Cardiol* 1991; **68**: 766–72.
- Girvin BG, Johnston GD. Comparison of the effects of a 7-day period of noncompliance on blood pressure control using three different antihypertensive agents. *J Hypertens* 2004; **22**: 1409–14.
- Heerdink ER, Urquhart J, Leufkens HG. Changes in prescribed drug dose after market introduction. *Pharmacoepidemiol Drug Saf* 2002; **11**: 447–453.
- Hughes DA, Walley T. Predicting “real world” effectiveness by integrating adherence with pharmacodynamic modeling. *Clin Pharmacol Ther* 2003; **74**: 1–8.
- Pullar T, Kumar S, Tindall H, Feely M. Time to stop counting the tablets? *Clin Pharmacol Ther* 1989; **46**: 163–8.
- Rudd P, Byyny RL, Zachary V, LoVerde ME, Titus C, Mitchell WD, Marshall G. The natural history of medication compliance in a drug trial: limitations of pill counts. *Clin Pharmacol Ther* 1989; **46**: 169–76.
- Stone AA, Shiffman S, Schwartz JF, Broderick JF, Hufford MR. Patient noncompliance with paper diaries. *BMJ* 2002; **324**: 1193–4.
- Turner BJ, Hecht FM. Improving on a coin toss to predict patient adherence to medications. *Ann Intern Med* 2001; **134**: 1004–6.
- Urquhart J. The electronic medication event monitor – lessons for pharmaco-therapy. *Clin Pharmacokinet* 1997; **32**: 345–56.
- Urquhart J. The odds of the three nons when an aptly prescribed medicine isn’t working: noncompliance, nonabsorption, nonresponse. *Brit J Clin Pharmacol* 2002; **54**: 212–20.
- US Centers for Diseases Control. Ten great public health achievements – US, 1900–99: Family planning. *Morbid Mortal Weekly Rep* 1999; **48**: 1073–80.
- Vrijens B, Urquhart J. Impact of variable adherence on antiretroviral drug action. *J Antimicrob Chemother* 2005; **55**: 616–27.
- Vrijens B, Comte L, Tousset E, Urquhart J. Once-daily versus twice-daily regimens: which is best for HIV-infected patients? 6th International Workshop on Clinical Pharmacology of HIV Therapy, 28–30 April 2005, Québec, Abstract 3, Poster 1.3.
- Vrijens B, Tousset E, Rode R, Bertz R, Mayer S, Urquhart J. Successful projection of the time-course of drug concentration in plasma during a one-year period from electronically compiled dosing-time data used as input to individually parameterized pharmacokinetic models. *J Clin Pharmacol* 2005; **45**: 461–7.
- WHO Tuberculosis Fact Sheet, number 104, rev April 2005. <http://www.who.int/mediacentre/factsheets/fs104/en/>.

Novel Approaches to Pharmacoepidemiology Study Design and Statistical Analysis

Edited by:

SAMY SUISSA

Department of Epidemiology and Biostatistics, Department of Medicine, and McGill Pharmacoepidemiology
Research Unit, McGill University, Montreal, Canada.

INTRODUCTION

The past two decades have witnessed an explosion of methodological advances in the design and analysis of epidemiological studies. Several of these contributions have been fundamental to the field of epidemiology in general, thus transcending content area, such as cancer, cardiovascular, occupational, or infectious disease epidemiology, to name a few. Further methodological advances have, on the other hand, arisen specifically from questions posed by pharmacoepidemiology applications or simply found a niche in pharmacoepidemiology because of the distinct nature of the available (and unavailable) data in this field, as well as its specific needs. 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.

First, we present various strategies of sampling within a large cohort, as an alternative to analyzing the full cohort. These sampling schemes are crucial in pharmacoepidemiology, where cohorts are necessarily large and the related expense and time of data collection and analysis for every single member of the cohort can be prohibitive. The nested case-control and case-cohort techniques are discussed for both internal and external comparisons of adverse event rates. Second, we describe techniques of design and analysis for situations where only partial data are available on confounders, such as only for the cases and not the controls of a case-control study, and we briefly introduce the two-stage sampling technique. Third, we describe new designs that use within-subject comparisons to estimate the risk of acute adverse events associated with transient drug effects, namely the case-crossover and case-time-control designs. We also briefly describe methods based solely on prescription drug databases, namely prescription sequence analysis

and prescription sequence symmetry analysis, to assess the risk of a drug as well as the phenomenon of channelling of drugs. Fourth, we describe a problem in the analysis of data from cohort studies of drug effectiveness that attempt to emulate the randomized controlled trial design, namely that of immortal time bias.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Pharmacoepidemiology deals with several facets of drug research, including the utilization, benefits, and risks of drugs. The primary focus of pharmacoepidemiology, however, and the one that receives the greatest attention and interest, is that of assessing the risk of uncommon, at times latent, and often unexpected adverse conditions resulting from the use of medications. Four features of nonexperimental research methods, affecting the degree of uncertainty in this risk assessment process, have recently been the object of methodological development and are the subject of this chapter.

First, because of the rarity of the adverse conditions under study, source populations and study cohorts must be extremely large to permit the control of statistical uncertainty arising from random error. It is not unusual to require population or cohort sizes in the tens or even hundreds of thousands of subjects to identify a sufficiently large number of subjects with the adverse condition under study to yield stable results. For example, the Cancer Prevention Study II cohort used 1.2 million persons to assess the effect of aspirin use on the risk of colon cancer, while the Nurses' Health Study cohort used 121 700 subjects to study the effect of oral contraceptives on the risk of cardiovascular diseases. Despite the advent of more powerful computers, the fact that drug exposure often varies over time and involves multiple agents complicates the analysis of cohort data. Further, sometimes supplemental data are needed on confounders or other variables, which may not be easily obtained on such large numbers. Accordingly, the first part of this chapter deals with the efficient designs within such cohorts that have been devised and can be used effectively in pharmacoepidemiology, providing accurate results more rapidly and at less expense.

The second source of uncertainty is related to the presence of confounding factors, which may possibly bias risk estimates and distort corresponding conclusions. For example, most epidemiologists accept unconditionally the finding that cervical clear-cell carcinoma in young women is caused by the use of diethylstilbestrol (DES) by their mothers

during pregnancy. Some, however, suggest that this is an unresolved issue, as the association may be confounded by the mother's history of spontaneous abortions and of bleeding during pregnancy. We therefore describe a novel approach based on confounder data measured only on cases, particularly suited to pharmacoepidemiology, as well as the two-stage sampling technique, which measures confounders on a sample of the case-control study population.

Third, pharmacoepidemiology is frequently faced with the assessment of the risk of rare acute adverse events resulting from transient drug effects. For example, to study the risk of ventricular tachycardia, from hypokalemia and prolonged Q-T intervals, associated with the use of inhaled β -agonists in asthma, the case-control approach may be challenging. The difficulty in selecting controls because of the acuteness of the adverse event, difficulties in determining the induction time and the timing of drug exposure, as well as possible confounding by indication may complicate this design. We thus describe the case-crossover and case-time-control design, devised to counter these complexities. The concept of comparing exposures within subjects used by these approaches also led to the development of several techniques applied to prescription drugs database studies.

Fourth, assessing drug effectiveness using an observation study design is particularly challenging, beyond the issue of confounding by indication. A recent such cohort study of over 22 000 elderly patients hospitalized for chronic obstructive pulmonary disease (COPD) suggested that inhaled corticosteroids given after discharge was associated with a 29% reduction in the rate of all-cause mortality, using a biased approach to data analysis. We will describe immortal time bias in the context of this study and address ways to counter this bias.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

SAMPLING WITHIN A COHORT

Cohort studies are essential to pharmacoepidemiology, as they form the basis for the quantification of drug risk assessment. By following users of the drug under investigation who are originally free of the adverse condition of concern until the occurrence of the adverse condition, a cohort enables the estimation of the rate of occurrence of this adverse event. Because of the usual rarity of the

adverse events under study, the cohort must be composed of very large numbers of subjects (see also Chapter 3). This type of requirement explains the infrequent use of the prospective cohort approach in pharmacoepidemiology. On the other hand, historical cohorts formed retrospectively from computerized databases have become a common and indispensable instrument in the armamentarium of many investigators in pharmacoepidemiology. Even with computerized cohorts, many studies may need to supplement and validate data obtained from the computer databases with data from hospital records, medical records, and physician or patient interview questionnaires. More importantly, however, drug exposures usually vary over time and include several drugs. The analysis of such a cohort with time-dependent exposure measures can be infeasible, if not impossible. To counter these constraints, designs based on sampling subjects within a cohort have been proposed and recently applied successfully in pharmacoepidemiology. We discuss structural aspects of cohorts and present two sampling designs within a cohort, the nested case-control and case-cohort designs, which permit the precise estimation of relative risk measures with negligible losses in precision.

PARTIAL CONFOUNDER DATA

The problem of confounding can be addressed, in the context of a case-control design, by either matching the controls to the cases with respect to all confounding factors, and to use a matched analysis, or to select unmatched controls and use statistical techniques to control for the effect of these confounders. Both approaches require the measurement of every confounding factor for each case and control subject. A situation encountered in pharmacoepidemiologic research is the availability of a wealth of data for the cases but a shortage of data for the controls. This is particularly true for studies based on administrative databases, where cases have likely been hospitalized and thus have an extensive medical chart, or died and have lengthy coroner or autopsy reports. For these cases, the investigator will have access to abundant information on potential confounding variables. However, if the controls are population based, they will not be able to provide these data on confounders. This chapter will describe a strategy of analysis to address this situation of partially available confounders, as well as the two-stage sampling approach for the situation when the confounders can be measured on only samples of the cases and controls.

WITHIN-SUBJECT DESIGNS

When conducting a case-control study, the controls should be selected to be representative of the source population that gave rise to the cases, to provide valid exposure information on the source population. This principle is often difficult to implement in practice, irrespective of whether one decides to select population or hospital controls. For population controls, we first can expect significant non-response or non-participation rates, which are problematic since the reasons for acceptance of participation as a control could be associated with exposure to the drug of interest, while the corresponding case selection is usually quite comprehensive. Second, when dealing with acute adverse events, the timing of the interview or data collection is crucial. For our example of the risk of ventricular tachycardia in association with the use of inhaled β -agonists in asthma, one would first select cases with this adverse event and easily probe whether they took the drug during the 4-hr span preceding the event. For controls, on the other hand, the investigator must define a time point of reference for which to ask the question about use of this drug in the "past 4 hrs." If, as a simple example, the drug is more likely to be required during the day, but controls can only be reached in the evening, the questioning process may become invalid since it will produce differential response patterns for cases and controls. For hospital controls, similar obstacles could invalidate the study. Consequently, when dealing with the study of transient drug effects on the risk of acute adverse events, the case-crossover design will be presented as a solution to these obstacles, along with the case-time-control design, which adjusts for exposure time trends.

IMMORTAL TIME BIAS

The analysis of cohort studies of drug effectiveness is a challenge if the drug exposure of interest changes over time. Several recent studies employed a time-fixed definition of exposure to emulate an intention-to-treat analysis used from randomized clinical trials. This principle is based on the premise that subjects are exposed to the drug under study immediately at the start of follow-up, which is not the case in observational database studies. For example, in the context of a COPD cohort study, subjects who received a prescription for inhaled corticosteroids during the 90-day period after cohort entry were considered exposed, while subjects who did not were considered unexposed. This approach, however, leads to immortal time bias, a major source of distortion in the rate ratio estimate. Immortal time refers to follow-up time during which, by definition

or design, outcome events cannot occur. Thus a subject dispensed their first prescription for an inhaled corticosteroid 80 days after cohort entry necessarily had to be alive on day 80, which generates an 80-day immortal time period. Thus, the exposed subjects will have a major survival advantage over their unexposed counterparts because they are guaranteed to survive at least until their drug was dispensed.

CURRENTLY AVAILABLE SOLUTIONS

SAMPLING WITHIN A COHORT

A cohort is defined by subjects meeting a set of eligibility criteria and by entry and exit time points. Consider, for example, the 13-year cohort study, spanned by the period 1978–1990, of the risks of human insulin in diabetes. For illustrative purposes, consider the subcohort of newly diagnosed diabetics only. Entry into the cohort may be defined by calendar time (spanned by the study, e.g., any time after January 1, 1978), by age (any age before 40th birthday), by events (the first use of a certain form of insulin), or by disease status (the date of diagnosis of diabetes). Exit from the cohort may be defined by the first occurrence of specific calendar time (e.g., December 31, 1990), age (exit at 40th birthday), events (death; exit from the study; the first use of an oral hypoglycemic agent), or disease status (onset of nephropathy).

Types and Structures of Cohorts

This cohort of newly diagnosed diabetics may be illustrated graphically as in Figure 26.1. This figure, based on 21 subjects, is plotted in terms of calendar time, with subjects ranked according to their date of entry into the cohort, which corresponds to disease diagnosis. Cohorts with this form of illustration, where the time axis of interest is calendar time (zero-time is January 1, 1978), depicting the chronological nature of the cohort, may be called *variable-entry cohorts*. An alternative depiction could be based on duration of disease (i.e., time since diagnosis or first exposure to insulin), which may be more relevant to the risk factor under study. In this instance, the illustration given in Figure 26.2 for the same cohort, using duration of disease as the new time axis, is significantly different from the previous one. Here, the subjects are ranked according to the length of follow-up time in the study and zero-time is the time of diagnosis. Such cohorts may be called *fixed-entry cohorts*. Alternatively, if a specific drug is of interest, zero-time can be redefined as the start of exposure to that drug, irrespective of when this occurs with respect to the time of disease diagnosis.

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, 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

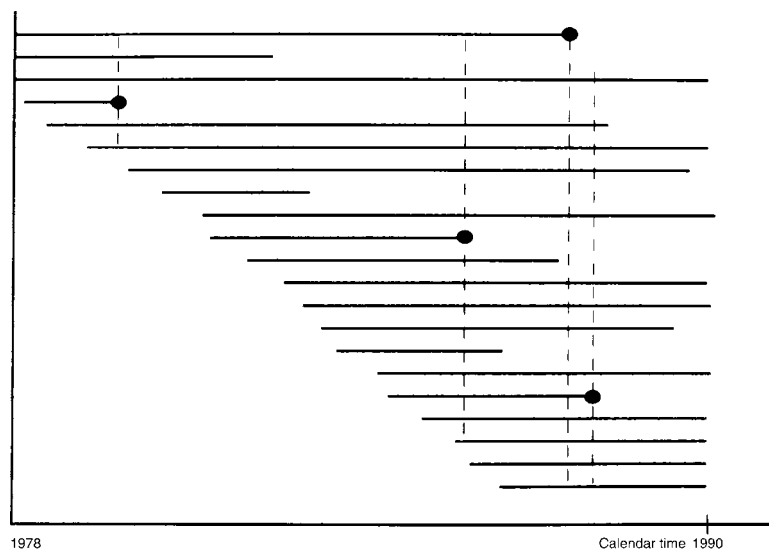


Figure 26.1. Illustration of a variable-entry cohort of 21 subjects followed from 1978 to 1990 with four cases (●) occurring and related risk-sets (---).

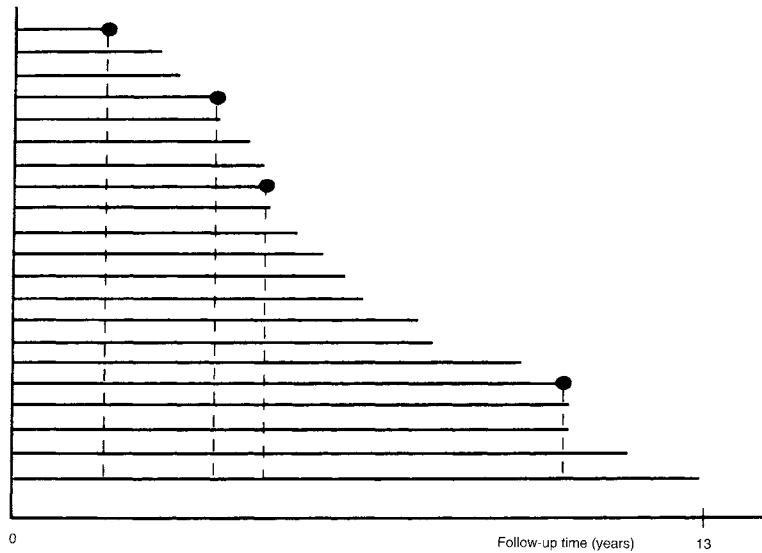


Figure 26.2. Illustration of fixed-entry cohort representation of the cohort in Figure 26.1, with new risk-sets (---) for the four cases.

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 adverse event at a given point in time, that is, they are free of the adverse event and are members of the cohort at that point in time. The only relevant risk sets for data analysis are those defined by the time of occurrence of each case. It is clear that Figures 26.1 and 26.2 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 26.1, for example, case 1 has in its risk set only the first 6 subjects to enter the cohort, while in Figure 26.2 all 21 cohort members belong to its risk set. In classical epidemiology, the second form (fixed-entry) based on disease duration is used almost exclusively in these situations, primarily because this time axis is the more important determinant of risk and exposure is assumed to be stable in time. In pharmacoepidemiology, on the other hand, drug exposure can vary substantially over calendar time, thus adding a “cohort effect.” Consequently, the first form (variable-entry) may be as relevant for the formation of risk sets and data analysis as the second form (see Case Example 26.1). Regardless, an advantage of having data on the full cohort is that we can change the primary time axis according to the question being posed, using calendar time for one analysis and duration of disease or drug exposure for another.

CASE EXAMPLE 26.1: **SAMPLING WITHIN A COHORT: NESTED** **CASE-CONTROL STUDY WITHIN A** **VARIABLE-ENTRY COHORT**

Background

- The incidence of *Clostridium difficile* infection associated with antibiotic use has increased over the last decade both in hospital and in community settings, despite a reduction in antibiotic use.

Question

- Could gastric acid suppressive drugs have contributed to the increase in the incidence of *Clostridium difficile* infection in the community?

Approach

- Conduct a nested case-control study within a population-based cohort of all subjects in the UK General Practice Research Database (GPRD).
- Compare each case of *Clostridium difficile* who had not been hospitalized in the prior year with 10 controls who also had not been hospitalized in the prior year, matched on follow-up time from entry into the practice and on calendar time.

(Continued)

Results

- The incidence of *Clostridium difficile* diagnosed in the community increased more than a 100-fold over the last decade.
- The calendar time specific rate of *Clostridium difficile* infection was increased with the use of gastric acid suppressive drugs, independently of antibiotic drug use.

Strengths

- Large stable population base with extensive computerized information.
- Calendar time matching essential to avoid confounding: *Clostridium difficile* incidence and gastric acid suppressive drug use increased considerably over time.
- Effective gastric acid suppression of these drugs enhances biological plausibility.

Limitations

- The rise in *Clostridium difficile* incidence may be due to increased testing, which is, however, not likely to be differential with respect to gastric acid suppressive drug prescription.
- While cases and controls had not been hospitalized in the prior year, it was not possible to assess whether they had been exposed to this bacteria in other ways.
- The case definition was based on a positive toxin assay and/or a clinical diagnosis. Sensitivity analysis on these two case definitions separately did not alter the results.

Summary Points

- Multiple drugs can be independently associated with the same adverse outcome.
- Matching on calendar time is crucial in studies where the exposure prevalence and outcome incidence both vary substantially over time, which is not uncommon in pharmacoepidemiology.

This “cohort effect,” important as a result of potentially significant drug exposure variation over calendar time, can be sufficiently accounted for by simply partitioning the cohort into several subcohorts, each having its own zero-time defined by entry date, and analyzing the duration within each subcohort. We could, for example, partition the cohort displayed in Figures 26.1 and 26.2 into four subcohorts, according roughly to 3-year intervals, to combine the two alternative forms of variable-entry and fixed-entry cohorts

illustrated in Figures 26.1 and 26.2. The risk sets from such a partition will necessarily depend on both disease duration and calendar time. Arguments for variable-entry cohorts can then be made by repeating the fixed-entry argument, conditional on each subcohort, and combining the results by stratification or regression methods. This would correspond to analyses controlling for calendar year. Because of the possibility of analyzing a variable-entry cohort as several fixed-entry subcohorts, we will focus the remaining presentation on a single fixed-entry cohort.

The Nested Case–Control Design

The idea of a nested case–control design within a cohort was first introduced by Mantel (1973), who proposed an unmatched selection of controls and called it a synthetic retrospective study. The nested case–control design involves four steps: defining the cohort’s time axis; selecting all cases in the cohort, i.e., all subjects with an adverse event; forming all risk sets corresponding to the cases; and finally randomly selecting one or more controls from each risk set. Figure 26.3 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, and that a subject may be selected as a control more than once. If, instead, 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, a bias is introduced in the estimation of the relative risk, because the control exposure prevalence will be slanted to that of longer-term subjects who do not become cases during the study follow-up. The magnitude of the bias depends on the frequency of the adverse event in the cohort; the more frequent the event, the larger the potential for bias.

This property leading to subjects possibly being selected more than once in the sample may be problematic when the exposure and covariate factors are time-dependent, particularly when the data are obtained by questionnaire; the respondent would have to answer questions regarding multiple time points in their history. This problem occurred in a study of the risks of severe adverse events in asthma associated with the use of inhaled β -agonists. A cohort of 12 301 asthmatics spanning the period 1978–87 was identified from the Saskatchewan Health computerized databases, of whom 129 were cases (death or near-death from asthma). To permit the collection of additional data from hospital charts and physician questionnaires, it was necessary to sample controls from the cohort. A standard nested case–control sample of six controls per case would have produced

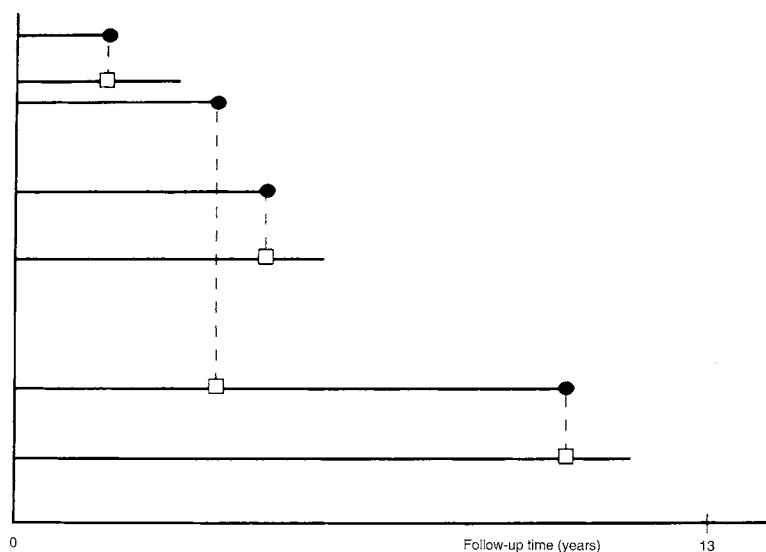


Figure 26.3. Nested case-control sample of one control □ per case (●) from the cohort in Figure 26.2.

some case and control subjects who contributed multiple times as controls in the sample. Filling out physician questionnaires on the same patient's asthma severity for different time points would be clearly confusing and problematic. To circumvent this difficulty, one can stratify the cohort according to various potential confounding factors, which in this study were age, area of residence, social assistance, prior asthma hospitalization, and calendar date of entry into the cohort. This extremely fine stratification resulted in 129 subcohorts, one for each case. 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, thus avoiding the problem of multiple time points.

The analysis of data from a nested case-control study must conserve the time-matched nature of the selection of cases and controls. Conditional logistic regression is the appropriate model as it uses the risk set as the fundamental unit of analysis, in agreement with the proportional hazards model of the full cohort. Computer packages such as EGRET, SAS or STATA can be used to fit this model.

The question of the required number of controls per case is important (see also Chapter 3). 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. It can be readily seen from 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. Although this general rule of an optimal 4:1 control-to-case ratio is appropriate in the majority of instances, one should be aware that when drug exposure is infrequent, when the hypothesized relative risk moves further from unity, or when several factors or other drugs are being assessed simultaneously, the ratio could easily be required to increase to 10 or more controls per case.

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 perform external comparisons, comparing the rate of adverse experience in the cohort to that of an external population, with appropriate adjustment for only a few available key factors, such as age, sex, and calendar time. The result is usually called the standardized mortality rate (SMR) or the standardized incidence rate (SIR). Since the nested case-control design is not a simple random sample from the cohort, a method to perform external comparisons using data from this design has been described.

The Case-Cohort Design

The first recognized application of a sampling design we currently call *case-cohort* was made by Hutchison (1968) in performing external comparisons of leukemia rates in patients treated by radiation for cervical cancer. It was ultimately developed and formalized by Prentice (1986), who coined the name "case-cohort." The case-cohort design

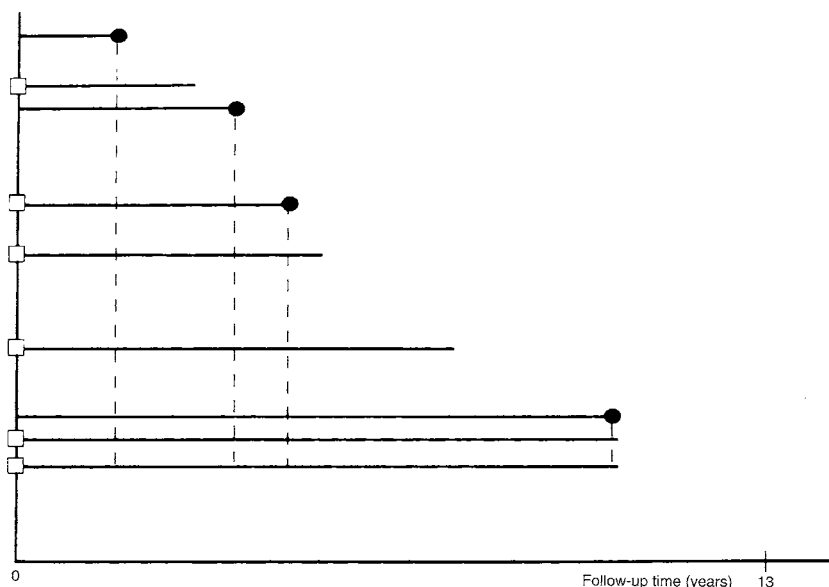


Figure 26.4. Case-cohort sample with six controls □ from the cohort in Figure 26.2.

involves two steps: selecting all cases in the cohort, i.e., all subjects with an adverse event, and randomly selecting a sample of predetermined size of subjects from the cohort.

Figure 26.4 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 added. It can also be perceived as an unmatched version of the nested case-control design. 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. The analysis for a case-cohort sample 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 has been released, which includes a module for the analysis of case-cohort data. As well, statisticians of the US National Cancer Institute are developing a user-friendly program called EPITOME, which performs this analysis. 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, e.g., its capacity to use the same sample to study different

outcomes, its ability to interchange the time axis of analysis from calendar to disease time, and its simplicity in performing external comparisons. Nevertheless, the nested case-control design is generally more practical, particularly since the analysis of case-cohort data can quickly become computationally infeasible with larger sample sizes and time-dependent exposures. 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.

PARTIAL CONFOUNDER DATA

A strategy of analysis exists, on the basis of data available solely for the cases of a case-control or a cohort study, to assess whether a factor is a confounder or not. The rationale is that, if the factor is not found to be a confounder in the cases by this method, the final analysis will not need to adjust for its effect when estimating the risk of the drug under study. The approach is described by Ray and Griffin (1989) and has been used in the context of a study of nonsteroidal antiinflammatory drugs and the risk of fatal peptic ulcer disease. The strategy is based on the definition of a confounder C ($C+$ and $C-$ denote presence and absence) in the assessment of the association between a drug exposure E ($E+$ and $E-$ denote exposure or not to the drug) and

an adverse condition D (D+ and D− denote presence and absence). Confounding is present if both of the following conditions are satisfied:

1. C and E are associated in the control group (in D−);
2. C and D are associated in E+ and in E−.

Because confounding assumes a common E : D odds ratio in C+ and in C−, condition 1 becomes equivalent to: C and E are associated in the case group (in D+). Thus, if in the cases we find no association between the potential confounder and drug exposure, confounding by this factor can be excluded outright, without having to verify the second condition. In this instance, the analysis involving drug exposure in cases and controls can be performed directly without any concern for the confounding variable. If, on the other hand, an association is found between C and E in the cases, confounding is not necessarily confirmed (since condition 2 must also be satisfied), but is very likely since a potential confounder is usually selected for its property of being a known risk factor for D.

Thus, if confounding is found to be present, the crude estimate is biased. A new method was recently developed to obtain an adjusted estimate of the rate ratio in the absence of confounder data among the controls. The adjusted odds ratio is given by:

$$OR_{adj} = P_0(w - y)/[(1 - P_0)y]$$

where $y = \{v - [v^2 - 4(r - 1)rxw]^{1/2}\}/[2(r - 1)]$, $v = 1 + (r - 1)(w + x)$ when $r \neq 1$ (and $y = wx$ when $r = 1$), r is the odds ratio between exposure and confounder among the cases, x is the probability of exposure among the controls, and w is the prevalence of the confounder among the controls. The latter (w) is the only unknown and must be estimated from external sources. An estimate of the variance of OR_{adj} exists. As an example, we use data from a case-control study conducted using the Saskatchewan computerized databases to assess whether theophylline, a drug used to treat asthma, increases the risk of acute cardiac death. In this study, the 30 cases provided data on theophylline use and smoking, a possible confounder, while the 4080 controls only had data available on theophylline use. Table 26.1 shows that the crude odds ratio between theophylline use and cardiac death is 4.3 [(17/13)/(956/3124)]. Among the cases, the odds ratio between theophylline use and smoking is estimable and found to be 7.5 [(14/5)/(3/8)], thus indicating that smoking is a strong confounder. Using an external estimate of smoking prevalence of 24% among asthmatics, obtained from a Canadian health survey, the adjusted odds

Table 26.1. Data from a case-control study of theophylline use and cardiac death in asthma, with the smoking confounder data missing for controls

	Cases		Controls	
	Use	Non-use	Use	Non-use
All subjects	17	13	956	3124
Stratified by smoking:				
Smokers	14	5	^a	^a
Non-smokers	3	8	^a	^a

^aThese frequencies are missing for controls.

ratio is 2.4, which is much lower than the crude estimate of 4.3, with 95% confidence interval of 1.0–5.8.

An alternative approach called two-stage sampling approach can be used when the confounders can be measured on a sample of cases and controls. Stage 1 involves the collection of information on drug exposure and outcomes, and stage 2 is the collection of confounder data on a subset of the stage 1 subjects. This situation, common with database studies, has recently been used in pharmacoepidemiology. An analogous method has been devised in the context of verifying the validity of the case status in studies where both outcome and exposure are rare.

WITHIN-SUBJECT DESIGNS

When dealing with the study of transient drug effects on the risk of acute adverse events, Maclure (1991) submits that the best representatives of the source population that produced the cases would be the cases themselves: this is the premise of the case-crossover design where comparisons between exposures are made within subjects. Other within-subject methods such as the case-time-control design and prescription symmetry analysis have been proposed and are also briefly presented here.

The Case-Crossover Design

A case-crossover 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 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. Also, the effect period (or time-window of effect) must be precisely determined. An incorrect specification of this time-window can have important repercussions on the risk estimate, as we

will show in the example below. Finally, 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.

The case-crossover study is simply a case-control study *in the cases only* arising from a crossover model. 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 because all are cases by definition. Each case is investigated to determine whether drug exposure was present or not within the predetermined 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 to the time-window of effect. This is done by obtaining data for a sufficiently stable period of time. 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 (1991).

Table 26.2 displays data generated from a hypothetical case-crossover study of 10 asthmatics who experienced ventricular tachycardia. The determination of drug exposure within the effect period for the “case” is straightforward.

Table 26.2. Hypothetical data for a case-crossover study of β-agonist exposure in the last 4 hours and the risk of ventricular tachycardia in asthma

Case no.	β-agonist use ^a in last 4 hours (a _i)	Usual β-agonist use in last year	Periods of risk (N _{1i})	Periods of no-risk (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

^aInhalations of 200 μg : 1 = yes, 0 = no. $RR = \sum a_i N_{0i} / \sum (1 - a_i) N_{1i}$.

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-hr periods being 2190 in one year. Thus, for example, the content of the 2 × 2 table for the first case, who is not found to have been exposed in the prior 4-hr 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 hrs, then the data of Table 26.2 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 hrs 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 hrs long, by considering as exposed cases 3 and 5, the relative risk becomes 5.0 (95% CI 2.0–12.2).

This method is 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, except if precise time-dependent data are available. The case-crossover design can be subject to selection bias if case selection is related to the exposure under study. Finally, information bias resulting from the differential quality of recent and past drug exposure data can be problematic if the exposure collection system is not robust, which is not a problem with computerized databases. This within-subject approach has also been adapted using a cohort approach for application to the risk assessment of vaccines. Case Example 26.2 provides an example of a case-crossover design to assess a potential adverse effect of vaccinations.

CASE EXAMPLE 26.2: CASE-CROSSOVER STUDY OF VACCINATION RISK IN MULTIPLE SCLEROSIS

Background

- The occurrence of relapses in multiple sclerosis is highly variable and unpredictable. Vaccines, particularly for hepatitis B, have been suspected to induce relapses in multiple sclerosis.

Question

- Does vaccination increase the rate of relapse in multiple sclerosis?

Approach

- A case-crossover study within the European Database for Multiple Sclerosis network. Cases with a relapse after a 12-month relapse-free period were questioned on vaccinations. Exposure to vaccination in the two-month risk-period immediately preceding the relapse was compared with that of the four previous two-month control periods to estimate the relative risk.

Results

- The prevalence of exposure during the two-month risk period was similar to that of the four control periods.
- The relative risk of relapse associated with exposure to any vaccination was thus unity.

Strengths

- Large clinical population with extensive computerized information.
- Efficient study design using only cases for this acute event and transient drug exposure.
- Confounding factors that do not change over time are inherently controlled for by the within-subject matched analysis.

Limitations

- Low vaccination prevalence in this clinical population does not permit assessment of the risk for shorter effect periods, such as a week.
- Confounding by factors that change over time, such as infections, could not be controlled for.

Summary Points

- Multiple sclerosis is highly variable over time and thus not easily amenable to cohort or case-control study designs.
- The case-crossover design is an efficient approach to study vaccine safety.

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* eliminates the biasing effect of unmeasured confounding factors, such as drug indication, while addressing the time trend assumption. In fact, the method is an extension of the case-crossover analysis that uses, in addition to the case series, a series of controls to adjust for exposure time trends.

Table 26.3 illustrates this approach with case-control data from the Saskatchewan Asthma Epidemiologic Project. The crude odds ratio for high β -agonist use is 4.4 (95% CI 2.9–6.7), while after adjustment for all available markers of severity the odds ratio is lowered 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 β -agonists was obtained for the one-year current period and the one-year reference period prior to the current period. Using a within-subject matched analysis among the 129 cases, the case-crossover odds ratio is 3.2 (29/9), while the control-crossover odds ratio among the 655 controls is 2.6 (65/25). 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).

Table 26.3. Illustration of a case-time-control analysis of data from case-control study of fatal or near-fatal asthma and β -agonist use

	Cases		Controls		Crude OR	Adjusted	
	High	Low	High	Low		OR	95% CI
Case-control analysis	93	36	241	414	4.4	3.1	1.8–5.4
Discordant ^a use (case-crossover)	29	9			3.2		1.5–6.8
Discordant ^a use (control-crossover)			65	25	2.6		1.6–4.1
Case-time-control	29	9	65	25	1.2		0.5–3.0

^a Discordant from exposure level during reference time period.

The case–time–control approach provides 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, 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, so caution is recommended in its use.

Drug Database Designs

Several computerized databases used in pharmacoepidemiology contain only information on prescriptions dispensed to patients, and no outcome information on disease diagnoses, hospitalizations, or vital status. These stand-alone prescription drug databases have been the object of recent methodological developments.

A technique that was developed specifically for the drug databases is prescription sequence analysis. Prescription sequence analysis is based on the situation when a certain drug A is suspected to cause an adverse event that itself is treated by a drug B. To apply this technique, the computerized drug database is searched for all patients who used drug A. For these subjects, all patients prescribed drug B in the course of using drug A are identified and counted. Under the null hypothesis that drug A does not cause the adverse event treated by drug B, the number of subjects should be proportional to the duration of use of drug A relative to the total period of observation. This technique was applied to assess whether using the anti-vertigo or anti-migraine drug flunarizine causes mental depression. An extension of prescription sequence analysis, called prescription sequence symmetry analysis, was recently proposed.

Another function of these databases is to use the prescriptions as covariate information to explain possible confounding patterns. The concept of *channeling* of drugs was put forward as an explanation of unusual risk findings. Conceptually the same as a selection bias due to uncontrolled confounding by indication (see Chapter 21), this phenomenon of channeling can be assessed rapidly in such databases, provided that medications can be used as proxies for disease severity. As this approach can be subject to bias when used with cross-sectional data, an application of channeling using a longitudinal design was recently presented.

IMMORTAL TIME BIAS

The analysis of cohort studies of drug effectiveness, where drug exposure is time varying, must necessarily be based

on time-dependent methods. In the example of the COPD cohort (see above), the “exposed” subjects who received their first prescription for an inhaled corticosteroid 80 days after cohort entry had to be alive on day 80. Thus, the subject was in fact “immortal” for the first 80 days and moreover was also unexposed, and must be classified accordingly. The combination of misclassification of unexposed person-time as exposed and the fact that this misclassified person-time is immortal will produce rate ratios lower than 1, thus creating an appearance of effectiveness for the drug.

Immortal time bias is thus the result of improper exposure definitions and analyses that cause serious misclassification of exposure and outcome events. This results from the attempt to emulate the randomized controlled trial to simplify the analysis of complex time-varying drug exposure data. However, observational studies rarely permit such simple paradigms. Instead, time-dependent methods for analyzing risks, such as the Cox proportional hazard models with time-dependent exposures or nested case–control designs (see above), must be used to account for complex changes in drug exposure and confounders over time.

THE FUTURE

The growing importance and awareness of pharmacoepidemiology in the medical, regulatory, and industry settings have led to a greater need and emphasis on solid methodology. As well, specific situations have induced the development of significant advances in the design and analysis of epidemiological studies of drug effects. We have described four recently developed methodologic approaches that facilitate the conduct of research in pharmacoepidemiology. Future developments in designs based on sampling within a large cohort, such as the nested case–control and case–cohort techniques, should provide user-friendly tools to facilitate the estimation of risk difference or excess risk measures and statistical models of analysis that take into account two or more time axes simultaneously. Methods dealing with partial confounder data could be directed to the situation of multiple confounders and effect modification, which eludes current techniques. Extensions and refinements of the case-crossover and case–time–control designs could address their assumptions as well as modifications for chronic effects and latent events. Lastly, the problem of immortal time bias reminds all pharmacoepidemiologists to remain vigilant with respect to the possible introduction of subtle yet important biases in the conduct of such research, as this problem appears to become more prevalent with the increasing use of database studies.

When used judiciously, these approaches can expand the limits inherent in the more traditional methods of epidemiology and generally optimize the conduct of research in pharmacoepidemiology. In the future, we can expect further enhancements of these methods and yet more effective tools in pharmacoepidemiology's unique search for the balance between high quality research and rapid results. This balance is fundamental to sound decision making around the management of drugs by clinicians, patients, industry, and regulators.

ACKNOWLEDGMENT

Samy Suissa is the recipient of a Distinguished Scientist award from the Canadian Institutes of Health Research (CIHR).

Key Points

- Pharmacoepidemiology studies must be efficient by offering rapid information of the utmost validity.
- Computerized databases provide valuable data sources for pharmacoepidemiological studies with unique methodological challenges.
- Epidemiological designs such as nested case-control and case-crossover are efficient approaches to assess the risks and benefits of drugs.
- Confounding bias can be assessed efficiently using subsets of the data.
- Immortal time bias can be avoided with proper time-dependent analyses.

SUGGESTED FURTHER READINGS

- Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of beta-agonists. *Am J Epidemiol* 1996; **144**: 1161–9.
- 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.
- Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol* 1996; **143**: 1165–73.
- Hallas J. Evidence of depression provoked by cardiovascular medication: a prescription sequence symmetry analysis. *Epidemiology* 1996; **7**: 478–84.
- 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.
- Hutchinson GB. Leukemia in patients with cancer of the cervix uteri treated with radiation. *J Natl Cancer Inst* 1968; **40**: 951–82.
- Lubin JH, Gail MH. Biased selection of controls for case-control analyses for cohort studies. *Biometrics* 1984; **40**: 63–75.
- 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.
- Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973; **29**: 479–86.
- Petri H, De Vet HCW, Naus J, Urquhart J. Prescription sequence analysis: a new and fast method for assessing certain adverse reactions of prescription drugs in large populations. *Stat Med* 1988; **7**: 1171–5.
- Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med* 1991; **10**: 577–81.
- Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; **73**: 1–11.
- Ray WA, Griffin MR. Use of Medicaid data for pharmacoepidemiology. *Am J Epidemiol* 1989; **129**: 837–49.
- Rothman KJ, Greenland S. *Modern Epidemiology*, 2nd edn. Hagerstown, MD: Lippincott-Raven, 1998.
- 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.
- Suissa S, Spitzer WO, Abenhaim L, Downey W, Gardiner RJ, Fitzgerald, D. *Risk of Death from Human Insulin*. New York: John Wiley & Sons, 1992; pp. 169–75.
- Suissa S. The case-time-control design. *Epidemiology* 1995; **6**: 248–53.
- Suissa S, Edwardes M. Adjusted odds ratios for case-control studies with missing confounder data in controls. *Epidemiology* 1997; **8**: 275–80.
- Suissa S, Edwardes MD, Boivin JF. External comparisons from nested case-control designs. *Epidemiology* 1998; **9**: 72–8.
- Suissa S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: immortal time bias in observational studies. *Am J Respir Crit Care Med* 2003; **168**: 49–53.
- Wacholder S, Boivin JF. External comparisons with the case-cohort design. *Am J Epidemiol* 1987; **126**: 1198–209.
- Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* 1991; **2**: 155–8.

SECTION IV

SPECIAL APPLICATIONS OF
PHARMACOEPIDEMIOLOGY

Special Applications of Pharmacoepidemiology

The following individuals contributed to editing sections of this chapter:

DAVID LEE,¹ SUMIT R. MAJUMDAR,² HELENE LEVENS LIPTON,³
STEPHEN B. SOUMERAI,⁴ SEAN HENNESSY,⁵ ROBERT L. DAVIS,⁶ ROBERT T. CHEN,⁷
ROSELIE A. BRIGHT,⁸ ALLEN A. MITCHELL,⁹ DAVID J. GRAHAM,¹⁰ DAVID W. BATES,¹¹
and BRIAN L. STROM⁵

¹Center for Pharmaceutical Management, Management Sciences for Health, Arlington, Virginia, USA; ²Department of Medicine, University of Alberta, Edmonton, Alberta, Canada; ³School of Medicine, University of California at San Francisco, San Francisco, California, USA; ⁴Harvard Medical School and Harvard Pilgrim Health Care, Boston, Massachusetts, USA; ⁵University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; ⁶Associate Professor, Pediatrics and Epidemiology, University of Washington, Seattle, Washington, USA; ^{*}Immunization Safety Office, Office of the Chief Science Officer, Centers for Disease Control and Prevention, Atlanta, Georgia, USA; ⁷Immunization Safety Branch, National Immunization Program, Centers for Disease Control and Prevention, Atlanta, Georgia, USA; ^{*}Global AIDS Program, Centers for Disease Control and Prevention, Atlanta, Georgia, USA; ⁸Center for Devices and Radiological Health, Food and Drug Administration, Rockville, Maryland, USA; ⁹Slone Epidemiology Center, Boston University Schools of Public Health and Medicine, Boston, Massachusetts, USA; ¹⁰Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; ¹¹Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts, USA.

In this chapter, we will present selected special applications of pharmacoepidemiology, which include: studies of drug utilization; evaluating and improving physician prescribing; drug utilization review; special methodologic

issues in pharmacoepidemiologic studies of vaccine safety; pharmacoepidemiologic studies of devices; studies of drug-induced birth defects; pharmacoepidemiology and risk management; the use of pharmacoepidemiology to study

* Current post

medication errors; and hospital pharmacoepidemiology. Again, we present this information using a standard format, focusing on clinical and methodological problems, examples of solutions, and perspectives on the future. Each application section then ends with a Case Example, and key points are listed at the end of the chapter.

STUDIES OF DRUG UTILIZATION

Drug utilization was 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.” Accordingly, studies of drug utilization include not only studies of the medical and nonmedical aspects influencing prescribing, dispensing, administration, and taking of medication, but also the effects of drug utilization at all levels of the health care system.

Drug utilization studies may be *quantitative* or *qualitative* (Table 27.1). In the former, the objective is to quantify the present state, developmental trends, and 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 result from 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 (see Chapters 7 and 8); 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.

Qualitative studies assess the appropriateness of drug utilization, usually by linking prescription data to the reasons for the drug prescribing. 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, and length of therapy, or others such as: failure to select a more effective or less hazardous drug if available, use of a fixed combination drug when only one of its components is justified, or use of a costly drug when a less costly equivalent drug is available. In North America, these studies are known as *drug utilization review* (DUR) or *drug use evaluation* (DUE) studies. For example, a large number of studies have documented the extent of inappropriate prescribing of drugs, particularly antibiotics, and the associated adverse clinical, ecological, and economic consequences.

DUR and DUE studies, aimed at problem detection and quantification, are usually one-time projects, not routinely conducted, 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 (Table 27.1). A DUR or DUE *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 (Table 27.1). Such programs are discussed in more detail in the “Drug Utilization Review” section of this chapter. The measurement of the effectiveness of the intervention is an integral part of the program.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

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—that is, for the right patient, with the right disease, in the proper dosage and intervals, and for the appropriate length of time. However, used

Table 27.1. 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

inappropriately, drugs 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.

Adverse drug reactions and drug noncompliance are important causes of preventable adult and pediatric hospital admissions. 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 toxicity risk would be 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. Medication error (discussed in the "Medication Errors" section of this chapter), poor patient compliance (Chapter 25), discontinuation of therapy, and problems in communication resulting from modern-day fragmentation of patient care may contribute to increased morbidity and mortality.

Therapeutic practice, as recommended, is based predominantly on data available from premarketing clinical trials. Complementary data from studies in the postmarketing period are needed to provide an adequate basis for improving drug therapy. Regardless, drug utilization studies address the relationship between therapeutic practice as recommended and actual clinical practice.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Considerable drug use data may be obtainable or are already available, the usefulness of which depends on the purpose of the study at hand. All have certain limitations in their direct clinical relevance. 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 data available are only approximations of this, 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. Again, both the available exposure and diagnosis data are suboptimal. Also, the criteria to be used to define "appropriate" are arbitrary.

Since most statistics on drug consumption were compiled for administrative or commercial reasons, the data are usually expressed in terms of cost or volume. 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.

Volume data are also available, as the overall weight of the drug sold or the unit volume sold (i.e., the number of tablets, capsules, or doses sold). This is closer to the number of patients exposed. However, tablet sizes vary, making it difficult to translate weight into the number of tablets. Prescription sizes also vary, so it is difficult to translate 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 latter is, of course, 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 drugs in each prescription. Finally, it should be noted that all these units represent approximate estimates of true consumption. The latter is ultimately modified by the patients' actual drug intake (i.e., degree of compliance; Chapter 25).

From a quality of care perspective, to interpret drug utilization data appropriately it is necessary 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.

EXAMPLES OF CURRENTLY AVAILABLE SOLUTIONS

Current Data Sources

Because of the importance of studying drug utilization, many computer databases have been used for DUR studies (Table 27.2). Most of these data sources lack information on

Table 27.2. Some computer databases for drug utilization studies

Not diagnosis-linked	Diagnosis-linked
<i>North America</i>	
National Prescription Audit ^a	National Disease and Therapeutic Index ^a
US Pharmaceutical Market—Drugstores ^a	Kaiser Permanente Medical Plan ^a
US Pharmaceutical Market—Hospitals ^a	Group Health Cooperative ^b
Medicaid Management Information Systems	The Slone Survey ^c
Saskatchewan Health Plan ^b	
<i>Europe</i>	
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	
Spain's Drug Data Bank (National Institute of Health)	
Denmark's Odense Pharmacoepidemiologic Database	
Denmark's County of North Jutland Pharmacoepidemiologic Prescription Database	

^aIMS America, Ltd.
^bPatient-specific data available for longitudinal studies.
^cReason for use.

morbidity and are mostly used for generating drug statistics and descriptive studies of drug consumption patterns. Some collect data in the form of drug sales, drug movement at various levels of the drug distribution channel, pharmaceutical or medical billing data, or all prescriptions dispensed.

Although the use of health insurance databases has also been reported in countries outside North America and Europe, medical and pharmaceutical databases are generally not available in most developing countries. An indicator-based approach, developed in the early 1990s by the International Network for Rational Use of Drugs (INRUD) and WHO, has facilitated the study of drug utilization in developing countries. It includes recommendations on core and complementary indicators, minimum sample sizes, sampling methods, and data collection techniques, depending on study objectives.

Units of Measurement

The defined daily dose (DDD) methodology was developed in response to the need to convert and standardize readily available volume data from sales statistics or pharmacy inventory data into medically meaningful units, to make crude estimates of the number of persons exposed to a particular medicine or class of medicines. 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 DDD methodology is useful for working with readily available gross drug statistics and is relatively easy and inexpensive to use. However, the DDD methodology 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, DDDs may be high or low relative to actual prescribed doses. 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, the concept and its applicability have not been incorporated into the WHO methodology. Finally, DDDs do not, of course, take into account variations in compliance.

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. Problems may arise in calculating the PDD due to a lack of clear and exact dosage indication in the prescription and dosage alteration via verbal instructions between prescribing events. 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. Although the DDD and the PDD may be used to estimate population drug exposure "therapeutic intensity," the methodology 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

Classification systems are used to categorize drugs into standardized groups. For example, the Anatomic Therapeutic Chemical (ATC) classification system is generally used in conjunction with the DDD methodology. The ATC system consists of five hierarchical levels: a main anatomical group, two therapeutic subgroups, a chemical-therapeutic subgroup, and a chemical substance subgroup. Medicinal products are classified according to the main therapeutic indication for the principal active ingredient. Use of the ATC classification system is recommended for reporting drug consumption statistics and conducting comparative drug utilization research. The WHO International Drug Monitoring Program uses the system for drug coding in adverse drug reaction monitoring and some developing countries have begun to use the ATC system to classify their essential drugs, which may eventually lead to preparation of drug utilization statistics.

The US uses the Iowa Drug Information System (IDIS), which is a hierarchical drug coding system based on the three therapeutic categories of the American Hospital Formulary Society (AHFS), to which a fourth level was added to code individual drug ingredients. Other coding systems, such as the National Drug Code and the Veterans' Administration Classification, do not provide unique codes for drug ingredients.

THE FUTURE

From a public health perspective, the observed differences in national and international patterns of drug utilization

require much further study. The medical consequences and the explanations for such differences are still not well documented. Analysis of medicine use by gender and age group may suggest important associations, as in a recent study on antidepressant medication use and decreased suicide rates. The increasing availability of population-based data resources will facilitate studies of incidence and prevalence of medicine use by age and gender.

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. Although regulation is effective, it is not possible to regulate all aspects of the clinical decision-making process to ensure optimal drug prescribing.

Many strategies aimed at modifying prescribing behavior have been proposed and adopted. Current evidence indicates that mailed educational materials alone are not sufficient to modify prescribing behavior. For interventions shown to be effective in improving drug prescribing, 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 the outpatient and inpatient settings, and the generalizability of face-to-face methods.

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. Patient outcome measures and process measures of quality of drug utilization have to be included in such studies.

In summary, the study of drug utilization continues to evolve. The development of large computerized databases that allow the linkage of drug utilization data to diagnoses, albeit subject to some inherent limitations, is contributing to this field of study. The WHO/INRUD indicator-based approach to drug utilization studies facilitates the development of drug utilization research in developing countries. Drug utilization review programs, particularly approaches that take into primary consideration patient outcome measures, merit further rigorous study. Opportunities for the study of drug utilization continue to be virtually unexplored, but the political issue regarding the confidentiality of medical records, as well as the shortage of funds and manpower in the current era of cost-containment, will determine the pace of growth of drug utilization research.

CASE EXAMPLE 27.1:
ASSESSING PRESCRIBER ADHERENCE TO
TYPE 2 DIABETES TREATMENT GUIDELINES

Background

- Prescriber adherence to treatment guidelines and patient adherence to prescribed treatment are important for metabolic control in diabetes type 2 patients. Identification of treatment adherence problems is important for designing interventions to improve medication use.

Question

- Is medication use consistent with current standards of care and what is the resulting degree of metabolic control?

Approach

- Prospective patient identification and selection.
- All patients attending follow-up visit over a five-month period.
- Medical record review and patient interview with structured questionnaire.

Results

- Very high proportion of patients with deficient metabolic control (40%).
- In an important proportion of patients (19%), there was disagreement between patient reported dose of antidiabetic medication and the dose in the medical record.
- In less than half of the patients, the treatment in the medical record was in agreement with published recommendations.

Strengths

- Prospective approach for identification of eligible study population.
- Assessed agreement of prescribed therapy with published standards.
- Assessed outcomes (metabolic control) of prescribed therapy.

Limitations

- Method for assessing patient adherence was lacking, although its importance was discussed, and is the link between prescribing and outcome.
- Only 59% of eligible patients were included in the study; most were not because they did not attend the facility on the scheduled day.
- Could not explain cause of disagreement between medical record dose and patient-reported dose.
- Quality of medical record notes may have contributed to study findings.
- Results not generalizable.

Summary Points

- Findings are useful for problem identification and appropriate follow-up.
- Good or acceptable metabolic control is not assured solely on provider adherence to published standards of care.
- Intervention should be multifaceted, aiming at provider and patient.

**EVALUATING AND IMPROVING PHYSICIAN
 PRESCRIBING**

If physicians and other health practitioners fail to update their knowledge and practice in response to new evidence about the outcomes of specific prescribing patterns, then pharmacoepidemiologic research may have little impact on clinical practice. Thus, the science of assessing and improving prescribing has grown rapidly; much of the growth 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 innovation diffusion, persuasive communications, adult learning theory, and knowledge translation.

**CLINICAL PROBLEMS TO BE ADDRESSED BY
 PHARMACOEPIDEMIOLOGIC RESEARCH**

Issues related to underuse, overuse, and misuse of medications all belong in the domain of prescribing “errors.” Factors responsible for suboptimal prescribing include:

failure of clinicians to keep abreast of new evidence; excessive promotion by the pharmaceutical industry; errors of omission; patient demands; and clinical inertia. Such diverse influences suggest the need for tailoring intervention strategies to the key factors affecting any given behavior based on models of knowledge translation.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Internal Validity

Poorly controlled studies produce misleading estimates of effect (Figure 27.1). Many nonintervention factors can affect medication use over time. Indeed, the “success” of many uncontrolled studies is often due to the attribution of pre-existing secular or temporal trends in practice to the intervention under study. Because randomized controlled trials (RCTs) are sometimes not feasible (e.g., contamination of controls within a single institution) or ethical (e.g., withholding quality assurance programs from controls), other strong quasi-experimental designs (e.g., interrupted time-series with or without comparison series, pre-post with concurrent comparison group studies) should be used instead of relying on weak one-group post-only or pre-post designs that do not generally permit causal inferences.

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. This argues once again for the need to conduct RCTs and well-controlled quasi-experiments to establish the effectiveness of interventions before they become a routine part of quality improvement programs.

Unit of Analysis

A common methodological problem in studies of physician behavior is the incorrect use of the patient as the unit of analysis. This violates basic statistical assumptions of independence because prescribing behaviors for individual patients are likely to be correlated within an individual physician’s practice. Such hierarchical “nesting” or statistical “clustering” often leads to accurate point estimates of effect but exaggerated significance levels and inappropriately narrow confidence intervals when the correct unit of analysis ought to have been the physician or practice or facility. Consequently, interventions may appear to lead to “statistically significant” improvements in prescribing when no such claim is warranted. Fortunately, methods for analyzing clustered data are available that can simultaneously control for clustering of observations at the patient, physician, and facility levels.

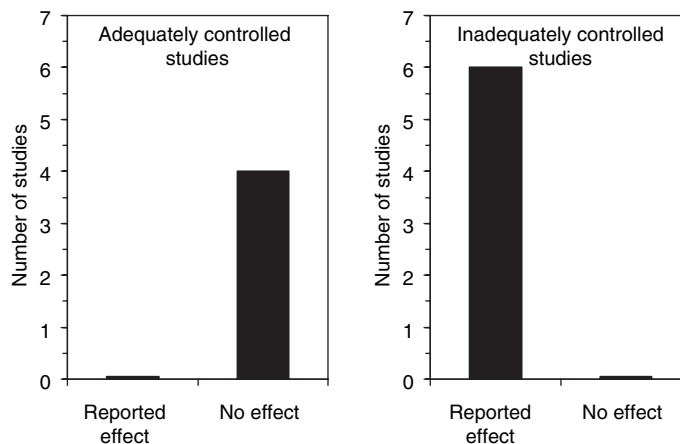


Figure 27.1. Reported effectiveness of dissemination of printed educational materials alone in well-designed versus inadequately controlled studies. Reprinted with permission from the *Milbank Quarterly* (Soumerai et al., 1989).

Ethical and Legal Problems Hindering the Implementation of Randomized Clinical Trials

It has been argued that there are ethical and legal problems related to “withholding” interventions designed to improve prescribing. This explicitly assumes that the proposed interventions will be beneficial. In fact, the effectiveness of many interventions is the very question that should be under investigation. Some have argued that mandating interventions without adequate proof of benefit is unethical. What is important is to demonstrate that such interventions are safe, efficacious, and cost-effective *before* widespread adoption.

Detecting Effects on Patient Outcomes

Few large well-controlled studies have linked changes in prescribing to improved patient outcomes, i.e., a link between improvements in process and patient outcomes such as mortality, for example. Sample sizes may need to be enormous to detect even modest changes in patient outcomes. 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.

EXAMPLES OF 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 nonclinical factors that could help or impede desired changes in clinical behaviors. One such model—PRECEDE—was developed for adult health education programs by Green and Kreuter (1991), and proposes factors influencing three sequential stages of behavior change: predisposing, enabling, and reinforcing factors. *Predisposing* variables include such factors as awareness of a guideline, knowledge of the underlying evidence, or beliefs in the efficacy of treatment. However, while a mailed drug bulletin may predispose some physicians, 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, or positive feedback) may be necessary to establish fully the new behavior. Such a framework explains a common observation: namely, that 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

There are numerous research syntheses that have evaluated the effectiveness of the most commonly studied interventions, including: dissemination of educational materials and guidelines; group education; profiling, audit, and feedback; reminders and computerized decision support systems; opinion leaders or educationally influential physicians; face-to-face educational outreach; financial incentives and penalties; and so on.

Distributing printed educational materials aimed at improving prescribing practice remains the most ubiquitous form of prescribing education. Unfortunately, use of disseminated educational materials alone may affect some predisposing variables (e.g., knowledge or attitudes) but have little or no effect on actual prescribing practice. Group education methods traditionally involve large-group didactic continuing medical education, but they are less effective than small group discussions conducted by clinical leaders.

Another popular approach to improving physician performance is providing physicians with clinical feedback regarding their prescribing practices, either in the form of comparative practice patterns with peers or predetermined standards such as practice guidelines, or in the form of patient-level medication profiles intended to highlight excessive, duplicative, or interacting medication use. Most types of feedback have a minimal effect on prescribing and medication profiles ought to be abandoned. On the other hand, computerized reminders enable physicians to reduce errors of omission by issuing alerts to perform specific actions in response to patient-level information such as laboratory findings or diagnoses, but excessive reminders may create “reminder-fatigue.” In addition, while studies of reminders have generally been positive, this does not extend to rigorously examined computerized decision support systems that attempt to move beyond the “secretarial” function of reminders.

Identifying local opinion leaders is another approach to help in the adoption of new pharmacological agents. In addition to opinion-leader involvement, this approach includes 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. Academic detailing

programs combine professionally illustrated materials with brief face-to-face visits by university-based pharmacists or peers, and have been consistently successful in reducing prescribing of contraindicated or marginally effective therapy. What sets academic detailing apart from industry detailing is that the messengers and the messages of the former are independent, objective, and based on the totality of the available evidence.

Finally, using financial incentives has been shown by numerous observational studies to affect the way that physicians practice medicine, and to be more powerful than using penalties—unfortunately, there have been no studies confirming that such methods are safe and without unintended consequences.

THE FUTURE

In general, it is clear that long-term changes in practice depend on inclusion of multiple strategies that predispose, enable, and reinforce desired behaviors. The following characteristics recur in successful interventions:

- identifying key factors influencing prescribing decisions through surveys, focus groups, or in-depth interviews;
- targeting physicians in need of education (e.g., review of prescribing data);
- 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 to be extinguished;
- brief educational materials to predispose and reinforce messages;
- use of multiple strategies to address multiple barriers to best practice;
- an 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 various broader questions still need to be answered, including issues related to opportunity costs and cost-effectiveness.

Currently, we know that prescribing problems exist, but we know little about their prevalence or determinants. This paucity of data is all the more remarkable considering that three-quarters of all physician visits end in the prescription of a drug. Future research efforts need to describe in greater detail the nature, prevalence, rate of prescribing, and severity of prescribing problems associated with the overuse, misuse, and underuse of medications (as previously discussed in the Drug Utilization section). Finally, studies examining the relationship between interventions and clinical outcomes would advance the field. 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 in its own right, it remains a critically important but relatively neglected area for rigorous study and investigation.

CASE EXAMPLE 27.2: **TESTING, RATHER THAN ASSUMING, THE** **EFFECTIVENESS OF AN INTERVENTION** **TO IMPROVE PRESCRIBING**

Background

- Because many experts assume that computerized decision support will be the magic bullet that improves physician prescribing, vast resources are being committed to widespread implementation before rigorous evaluation.

Question

- Can computerized decision support improve the quality of ambulatory care for chronic conditions such as ischemic heart disease or asthma?

Approach

- Eccles and colleagues (2002) conducted a cluster randomized controlled trial of 60 primary care practices in the United Kingdom; 30 practices randomized

(Continued)

to heart disease and 30 to asthma, with each practice acting as a control for the non-assigned condition. Practices already had computerized health records and many had electronic prescribing.

- Compared about 40 measures of guideline adherence (prescribing, testing, patient reported outcomes) between intervention and control practices one year after implementation of decision support.

Results

- No effect of computerized decision support on any measure of guideline adherence for either heart disease or asthma.
- Hypothesized reason for lack of effect was extremely low levels of use of, and dissatisfaction with, the decision support tools by physicians.

Strengths

- Cluster randomized trial design eliminated selection and volunteer bias, while controlling for any secular improvements in quality of care.
- Overcomes almost all of the flaws in design and analysis of previous studies of computerized decision support.
- Conducted in a “real world” setting rather than with housestaff or within academic practices or in the hospital setting.

Limitations

- Newer technologies or better delivery systems may be more acceptable to primary care physicians.
- Inadequate attention may have been paid to the non-technological factors that influence the acceptability and use of an intervention (e.g., lack of up-front buy in from end-users, lack of incentives for using the system, lack of participation in the actual guideline development process).

Summary Points

- Just as with any drug or device, interventions to improve prescribing should be tested in controlled studies before widespread adoption.
- Results of interventions (including decision support) conducted at specialized academic centers or in the hospital setting may not necessarily be applicable to the “real world” of busy primary care practice.

DRUG UTILIZATION REVIEW

Drug utilization review (DUR) programs have been defined as “structured, ongoing initiatives that interpret patterns of drug use in relation to predetermined criteria, and attempt to prevent or minimize inappropriate prescribing.” DUR has many synonyms, including *drug use review*, *drug use evaluation*, and *medication use evaluation*. As discussed earlier in this chapter, DUR programs differ from *drug utilization studies*, which are time-limited investigations that measure drug use, but do not necessarily assess appropriateness or attempt to change practice.

DUR programs are categorized by the timing of interventions within the drug use process, with *prospective* DUR occurring before the patient receives the medication and *retrospective* DUR occurring after the patient has received the medication. Recently, the use of clinical decision support within computerized prescriber order entry (CPOE) programs has risen dramatically. The use of such programs to improve prescribing can be considered a form of prospective DUR in which prescribers are the targets of interventions, and was discussed previously in this chapter.

Generally, the DUR process involves comparing actual behavior to explicit, prospectively established standards, referred to as *criteria*. For example, a commonly used criterion is that patients should not receive more than one nonsteroidal anti-inflammatory agent at any one time. Criteria have been developed to identify the following types of problems: drug–drug interactions, drug–disease interactions, drug–age interactions, drug–allergy interactions, use of too high or too low a dose, duplication of therapeutic class, excessive duration of therapy, obtaining prescription refills sooner or later than should be needed, failure to prescribe a known effective agent in patients with certain conditions, abuse of psychoactive medications, and use of a more costly agent when a less costly agent is available.

After developing criteria, the next step in the DUR process is to measure adherence to explicit criteria by examining individual-level data. Instances in which medication use does not agree with criteria are called *exceptions*.

Next, interventions are implemented where appropriate, often following an implicit review. Although the general model for DUR does not require that practitioners be made aware of individual exceptions occurring in their patients (that is, interventions can be made based on aggregate rather than individual findings), this step usually involves alerting the physician and/or pharmacy of record as to the occurrence of the specific exception.

There are different settings in which the DUR model is applied. *Outpatient retrospective DUR programs* use computerized administrative data (i.e., pharmacy and medical claims data maintained for billing and other administrative purposes) to identify exceptions that are then reviewed by a physician or pharmacist, or by a committee of health professionals, and result in an intervention (e.g., a mailed alert letter to the physician). The alert letter typically describes the DUR program and the criterion, and provides literature references supporting the criterion and a patient profile demonstrating that the criterion was violated.

Outpatient prospective DUR programs include efforts that occur before or during the prescribing process, or during the dispensing process. Computer alerts are generated either by the pharmacy's own computer or the computer of the pharmaceutical benefit management company that reimburses the pharmacy for dispensing the prescription and are conveyed to the pharmacist filling the prescription. The responsibility of resolving the potential problem belongs to the pharmacist, who must contact the prescriber if the prescription is to be changed. The fact that the alert is conveyed to the pharmacist rather than the physician would be expected to reduce the effectiveness of pharmacy-based prospective DUR programs relative to physician-based interventions. Computer-aided review of patient profiles may have advantages over manual review by virtue of increased sensitivity in detecting potential problems. However, computer review systems based on overly inclusive criteria for producing alerts engender numerous apparently trivial alerts, and may thus foster disregard of alerts, reducing the effectiveness of such systems.

Hospital DUR programs are usually conducted by the pharmacy department acting in conjunction with and by the authority of a medical staff committee, such as the pharmacy and therapeutics ("P&T") committee. Hospital DUR programs use the same overall process as described above, although, unlike most outpatient DUR programs, they generally have access to primary medical records. Hospital DUR programs also tend to perform series of discrete evaluations rather than ongoing evaluations, and often include elements of both prospective and retrospective review.

FRAMEWORK FOR EVALUATING THE EFFECTIVENESS OF DUR

Real-world DUR programs operate by performing individual drug use audits, which are either discrete or ongoing. Thus, each criterion or small group of related criteria used

by a DUR program can be considered to be a distinct drug use audit. Studies focusing on the effects of specific drug use audits (i.e., the application of an individual criterion or small group of related criteria) are more specific, and thus more credible than evaluations of overall DUR programs.

Similarly, evidence for more effects on process measures, such as changes in prescribing, tends to be more credible than evidence for clinical outcomes such as reductions in the rate of all-cause hospital admission. This is because there is greater potential for external factors to influence the rate of clinical outcomes and because clinical outcomes are rarer.

While the entire DUR process can be seen as an application of pharmacoepidemiology *practice*, the primary role for pharmacoepidemiologic *research* in the area of DUR is to evaluate the effectiveness of these programs in meeting their stated goals. Clarifying the mechanisms by which DUR acts is also important.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

The ability to evaluate the effectiveness of DUR programs by means of randomized experiments is limited. One barrier in the US is that Medicaid programs are required by federal law to perform DUR, making randomized trials of entire DUR programs impossible within Medicaid. This would not, however, preclude randomized trials of specific components of DUR programs.

Another problem is that when randomized experiments have been performed, the need to keep the intervention and control groups free from cross-contamination has necessitated the randomization of *cluster* (e.g., units such as physicians, pharmacies, or geographic regions) rather than the randomization of individual patients. Randomization by cluster results in fewer randomized units and reduced statistical power.

An additional challenge to research in this area is that the serious clinical outcomes that DUR programs are intended to prevent (for example, serious gastrointestinal bleeding caused by NSAIDs) are relatively rare on a population basis. Therefore, studies attempting to show clinical effects of DUR programs need to be enormous in size.

Finally, the validity of any nonrandomized study is subject to threat by many factors, the most prominent of which are noncomparable study groups and unmeasured baseline differences between treated and untreated subjects.

EXAMPLES OF CURRENTLY AVAILABLE SOLUTIONS

Although there is evidence that some forms of retrospective DUR can modestly impact prescribing, there is no evidence that it achieves its primary objective of improving clinical outcomes, and some evidence that it does not. Further, there is some evidence suggesting that even thoughtfully considered and well-intentioned interventions aimed at improving care might have unintended consequences that need to be uncovered through rigorous research. Prospective DUR, while intuitively appealing, lacks evidence that it provides an incremental benefit over standard pharmacy practice. Although the conduct of rigorous research in this area is challenging, continued support for these programs as a means to improve clinical care should be based on empiric evidence. The only available evidence suggests the absence of an observable effect on clinical outcomes.

There are several plausible explanations for the apparent lack of effect of DUR programs as currently implemented. One is the unknown validity of many DUR criteria. For example, most information about drug–drug interactions comes from either case reports or controlled studies of effects on serum drug concentrations—a surrogate endpoint. Both types of evidence have limited utility for inferring causal effects on clinical outcomes, therefore there is a striking amount of disagreement among standard drug interaction references regarding which drug combinations need to be avoided. A second plausible explanation is that systems that provide too many alerts (especially those perceived to be baseless or clinically unimportant) result in user fatigue and disregard of alerts, and thus reduce the effectiveness of the systems. There is evidence for this in the setting of prospective DUR.

THE FUTURE

There is now a reasonable body of research suggesting that outpatient prospective and retrospective DUR programs have not produced a large or even measurable improvement in clinical outcomes on a population basis. However, unless there is a change in US federal law, DUR will continue to be used in Medicaid programs. Given the increasing use of CPOE programs that permit the incorporation of computerized decision support (including prospective DUR with alerts presented to the prescriber at the time of prescribing), we should expect the use of CPOE-based DUR to grow. However, given that major perceived flaws of existing DUR programs include criteria with questionable validity and unacceptably high alert rates, it would be naive to

expect that applying standard criteria to CPOE-based DUR would improve clinical outcomes. Indeed, prescribers are unlikely to be any more tolerant of apparently invalid alerts presented with high frequency than are pharmacists.

We believe that those implementing CPOE-based DUR should use a slow, incremental approach to building criteria, starting with apparently unambiguous criteria such as absolutely contraindicated therapies (e.g., drugs contraindicated in pregnant women). The effects of implementing specific criteria should be measured on appropriate endpoints. Also, the knowledge base underlying potential criteria needs to be strengthened before they are implemented. In addition, we need research into the best way to communicate alerts.

CASE EXAMPLE 27.3: **EVALUATION OF THE EFFECTIVENESS** **OF A DUR PROGRAM TO REDUCE** **UNNECESSARY USE OF HYPNOTICS**

Background

- Because of the risks associated with use of benzodiazepines as hypnotics, the State of Washington Medicaid program wished to measure the effectiveness of an intervention to reduce long-term use of specific benzodiazepines for insomnia.

Question

- Is the intervention effective in reducing use of targeted benzodiazepines, and are there any unintended consequences of the intervention?

Approach

- Cluster-randomized trial of physicians who saw at least one patient with prescription claims indicating that they had used at least one tablet per day of any target sedative hypnotic for at least one year. Physicians in the intervention group were mailed an alert letter; physicians in the control group received no intervention.

Results

- Patients seen by physicians in the intervention group received fewer prescriptions for target hypnotics than patients seen by physicians in the control group.
- However, despite recommendations in the alert letter to avoid abrupt discontinuation of benzodiazepines, some patients may have had therapy suddenly stopped.

- In addition, there was an *increased* use of non-targeted sleep agents associated with the intervention.

Strengths

- Randomized trial design eliminates confounding by indication.
- Contamination was avoided by randomizing physicians rather than individual patients.

Limitation

- Clinical effects were not observed.

Summary Points

- Strong study designs like randomized trials provide more convincing evidence for effect than weak designs like pre–post comparisons with no control group.
- Studies that randomize clusters need to account for clustering in the analysis.
- Quality improvement initiatives can have unintended consequences.

SPECIAL METHODOLOGIC ISSUES IN PHARMACOEPIDEMOLOGIC STUDIES OF VACCINE SAFETY

Vaccines are among the most cost-effective and effective public health interventions, and have led to considerable declines in vaccine-preventable diseases. No vaccine is perfectly safe or effective, however, and concerns over adverse events (AEs) after immunizations have often threatened the stability of immunization programs.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGIC RESEARCH

The tolerance for adverse reactions to vaccines given to healthy persons—especially healthy babies—is lower than to products administered to persons already sick. A higher standard of safety is required for vaccines because of the large number of persons who are exposed, some of whom are compelled to do so by law or regulation for public health reasons. These issues lead often to a need to investigate much rarer adverse events following vaccinations than for

pharmaceuticals. However, the cost and difficulty of studying events increase with their rarity, and it is difficult to provide definitive conclusions from epidemiologic studies of such rare events.

Due to the near-universal exposure to many vaccines, the maxim “first do no harm” is particularly cogent. These concerns are the basis for strict regulatory control and other oversight of vaccines by the FDA and the WHO. 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 choices, and the decision to withdraw a vaccine may have wide ramifications. Establishing associations of adverse events with vaccines and prompt definition of the attributable risk are critical in placing adverse events in the proper risk/benefit perspective. By virtue of vaccines being relatively universal exposures, despite the relative rarity of serious true vaccine reactions, the absolute number of clinically significant vaccine adverse event reports received annually in the US now averages ~15 000. Vaccine safety monitoring is a dynamic balancing of risks and benefits. When diseases are close to eradication, data on complications due to vaccine relative to that of disease may lead to discontinuation or decreased use of the vaccine.

Research in vaccine safety can help to distinguish true vaccine reactions from coincidental events, estimate their attributable risk, identify risk factors that may permit development of valid contraindications, and, if the pathophysiologic mechanism becomes known, develop safer vaccines.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGIC RESEARCH

An Institute of Medicine (IOM) review of vaccine safety found that the US knowledge and research capacity has been limited by: (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 to provide persuasive evidence of causation; and (v) few experimental studies published relative to the total number of epidemiologic studies published. In ensuing attempts to overcome these limitations, epidemiology has been vital in providing the scientific methodology for assessing the safety of vaccines.

Signal Detection

High profile vaccine adverse events, such as intussusception following Rotavirus vaccination, demonstrate the need for surveillance systems able to detect potential aberrations in a timely manner. Some factors make identification of true signals difficult, however; for example, many vaccines are administered early in life, at a time when the baseline risk is constantly evolving. Until the recent advent of systematic analyses of automated data and data mining of spontaneous reports methods, detection of a vaccine safety signal occurred as much due to a persistent patient as from data analysis.

Assessment of Causality

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), recurrence upon rechallenge (e.g., alopecia and hepatitis B vaccination), or a vaccine-specific laboratory finding (e.g., Urabe mumps vaccine virus isolation) can be identified. When the adverse event also occurs in the absence of vaccination (e.g., seizure), epidemiologic studies are necessary to assess whether vaccinated persons are at higher risk than unvaccinated persons.

Measurement Accuracy

Misclassification of exposure status may occur if there is poor documentation of vaccinations. Documentation of exposure status has been fairly good through school age, but difficulty has been encountered in ascertaining vaccination status in older persons.

Sample Size

Because outcome events (e.g., encephalopathy) are often extremely rare, it may be a challenge to identify enough cases for a meaningful study. The difficulty with adequate study power is further compounded in assessing rare events in populations less frequently exposed (e.g., subpopulations with special indications). For studies of rare outcomes, case-control designs are the most efficient. Such studies typically sample the source population of cases, identify an appropriate control group, and assess the exposure status of both groups to estimate the risk associated with exposure.

Confounding and Bias

Because childhood vaccines are generally administered on schedule and children may have developmental dispositions to particular events, age may confound exposure-outcome relations, e.g., DTP vaccine and febrile seizures or SIDS. Consequently, such factors must be controlled, generally by matching, as well as in the analysis.

More difficult to control are factors leading to delayed vaccination or nonvaccination. Such factors (e.g., low socioeconomic status) may confound studies of vaccine AEs and lead to underestimates of the true relative risks. Those who have not been vaccinated may differ substantially from the vaccinated population in risks of AEs and thus be unsuitable as a reference group in epidemiologic studies. The unvaccinated may be persons for whom vaccination is medically contraindicated, or they may have other risks (e.g., they may be members of low socioeconomic groups) for the outcome being studied.

EXAMPLES OF CURRENTLY AVAILABLE SOLUTIONS

Signal Detection

Identifying a potential new vaccine safety problem ("signal") requires a mix of clinical intuition and epidemiologic expertise. One tool for comparing safety profiles of vaccines is simple, and involves comparing the proportions of particular symptoms out of the total number of events for a given vaccine to that observed among reports for another vaccine or group of vaccines. Due to its ease of implementation and interpretation, this proportional reporting rate ratio (PRR) method is the most widely used measure in vaccine adverse event reporting systems (VAERS) for prospective and retrospective signal generation (see Chapter 8). Of course, such signals need to be confirmed in formal epidemiologic studies.

Epidemiologic Studies

Historically, *ad hoc* epidemiologic studies have been employed to assess potential vaccine adverse events. Newer projects based on automated large-linked databases provide a more flexible framework for hypothesis testing than *ad hoc* epidemiologic studies. To this end, the CDC initiated the Vaccine Safety Datalink (VSD) project in 1990. The VSD study prospectively collects vaccination, medical outcome (e.g., hospital discharge, outpatient visits, emergency room visits, and deaths), and covariate data (e.g., birth certificates, census) under joint protocol at multiple HMOs (see Chapter 12). The VSD focused its initial

efforts on examining potential associations between immunizations and a number of neurologic, allergic, hematologic, infectious, inflammatory, and metabolic conditions, and conducted surveillance on approximately 500 000 children from birth through six years of age. The VSD is also used to test new *ad hoc* vaccine safety hypotheses that may arise from the medical literature, VAERS, changes in immunization schedules, or introduction of new vaccines. Due to the high coverage attained in the HMOs for most vaccines, few non-vaccinated controls are available, and thus the VSD is limited in its capacity to assess associations between vaccination and adverse events with delayed or insidious onset (e.g., neurodevelopmental or behavioral outcomes).

The VSD provides an essential, powerful, and relatively cost-effective complement to ongoing evaluations of vaccine safety in the US, and similar systems have since been developed in the UK, Canada, and Vietnam.

THE FUTURE

As the infrastructure of the US immunization safety system has been strengthened, several nonscientific challenges have arisen that threaten this important research. The credibility of government-sponsored vaccine safety studies has been questioned by some, not so much because of the scientific merits, but based on perceived conflict of interest since the vaccine safety researchers are located within national immunization programs that “promote” immunizations. The credibility of researchers from the “other side” has also been questioned, due to use of faulty methodology and failure to disclose source of study funding from legal sources.

CASE EXAMPLE 27.4: ROTAVIRUS VACCINATION AND RISK OF INTUSSUSCEPTION

Background

- Intussusception was reported among recipients of a rhesus-human rotavirus reassortant-tetravalent vaccine (RRV-TV), RotaShield® (Wyeth Laboratories, Inc).

Question

- Is a vaccine containing four live viruses, a rhesus rotavirus (serotype 3) and three rhesus-human reassortant viruses (serotypes 1, 2, and 4) associated with intussusception in RRV-TV vaccine recipients?

Approach

- Analyze Vaccine Adverse Events Reporting System (VAERS) passive surveillance intussusception reports.
- Conduct a case-control and cohort study on RRV-TV recipients.
- Conduct postmarketing vaccine surveillance.
- Determine elevated risk associated with RRV-TV receipt and intussusception.

Results

- Fifteen initial VAERS reports (total of 112 VAERS intussusception reports from licensure on August 31, 1998 to December 31, 1999) with 95 confirmed intussusception cases following RRV-TV (confirmed by medical record review).
- Case-control study found infants receiving RRV-TV were 37 times more likely to have intussusception 3–7 days after the first dose than infants who did not receive RRV-TV (95% confidence interval [CI] = 12.6–110.1).
- Retrospective cohort study among 463 277 children in managed care organizations demonstrated that those receiving the vaccine, 56 253 infants, were 30 times more likely to have intussusception 3–7 days after the first dose than infants who did not receive the vaccine (95% CI = 8.8–104.9).
- Causal link between RRV-TV receipt and intussusception established in postmarketing period at a frequency detectable by current surveillance tools (approximately 1/5000–1/10 000 vaccinees).

Strengths

- Multiple studies confirmed association between RRV-TV and intussusception.
- Despite the limitations of passive surveillance, VAERS successfully provided a vaccine adverse event alert.

Limitations

- Passive surveillance systems such as VAERS are subject to multiple limitations, including underreporting, reporting of temporal associations or unconfirmed diagnoses, lack of denominator data, and unbiased comparison groups.

(Continued)

Summary Points

- The CDC recommended temporarily suspending use of RRV-TV following the initial 15 VAERS reports.
- No case of RRV-TV-associated intussusception occurred among infants vaccinated after the recommendation was issued on July 16, 1999.
- When the VAERS findings were substantiated by preliminary findings of the more definitive studies, the manufacturer voluntarily recalled the vaccine.
- The ACIP recommendations for RRV-TV vaccination were withdrawn in October 1999.
- A new “rapid cycle” initiative for more timely detection of vaccine safety signals has been formed by the CDC Vaccine Safety Datalink (VSD) project; this project successfully simulated and retrospectively “detected” the RRV-TV intussusception signal within the VSD by mid-May 1999.

PHARMACOEPIDEMIOLOGIC STUDIES OF DEVICES

The legal definition of “medical devices” as offered by the US Congress is:

The term “device”... means an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including any component, part, or accessory, which is recognized in the official National Formulary, or the United States Pharmacopeia, or any supplement to them, intended for the use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or intended to affect the structure or 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 its primary intended purposes.

Other countries use similar definitions.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Study Hypotheses for Epidemiologic Studies of Medical Devices

The main regulatory purpose of categorizing medical devices is to vary the level of information required before

marketing may begin, according to the potential risk posed by the particular device. The highest level of control is to require RCTs before applying for approval to market the device. This is similar to the expectations for a new drug. In contrast, if a new device is very similar to several that have already been cleared for marketing, then the device sponsor may need to demonstrate equivalence to the earlier devices, or attest that the device conforms to an international standard recognized by the regulatory agency. If the predicate device was never studied in an RCT (in the US, many devices marketed before the 1976 Medical Device Amendment have never been subjected to the requirement), gaps in human safety and efficacy data exist. Furthermore, the definition of “similar” is such that incremental changes are allowed; after an accumulation of such changes, the latest device may be quite different from the original predicate device. Further, a very large category of devices receives minimal regulation; the sponsor simply must register itself and the device with the agency and is expected to follow general guidelines. Devices may be moved to less regulated categories over time as clinical experience with its use expands, depending on the comfort of the agency with the regulatory history of the device.

Many regulatory agencies monitor the safety of devices by reviewing adverse event reports from users, sponsors, or the scientific literature. The US FDA also has a regulatory tool unique among members of the Global Harmonization Task Force: it may require the sponsor of a marketed device to conduct a “postmarketing surveillance study” of safety and/or effectiveness, if warranted by public health considerations.

Epidemiology study designs differ based on the use patterns of devices in terms of the number of patients exposed to a device, number of times a device is used, extent of direct patient exposure to the device, nature of device effect, permanence of a device, setting of device use, device user, etc. Distinct epidemiologic problems, in the sense of hypotheses to be tested and challenges to study validity, derive from each of these categorizations. Short-term safety and efficacy issues apply to all the device categories, except for diagnostics. Long-term safety and efficacy hypotheses apply to reused and durable equipment, and long-term implants. Human error issues related to proper use, and perhaps maintenance, pertain to all types of devices. The consequences of device reuse are worth studying for equipment (whether designed to be reused or durable, or reused in spite of being designed for single-use only). Finally, hypotheses regarding the validity of the test result are appropriate for diagnostic devices.

Public Health Impact

If epidemiologic evidence points to a safety problem with a device, further information is then required to evaluate the public health impact of doing nothing, taking an action, or perhaps taking an alternative action. Available epidemiologic evidence of the effectiveness and extent of device use, as well as the availability of alternative therapies, have a bearing on the decision.

Differential Effects

Epidemiology can help to identify patients at higher risk of complications from a device through the study of the differential effects of devices by some cofactor, such as gender or a concomitant therapy.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Recognition and Characterization of New Syndromes or Adverse Events

Sometimes the discovery of a device's role in causing a well-recognized outcome is difficult, partly because devices are generally taken to be benign. At other times, a new outcome must be defined. Epidemiologists can contribute to improved methodology to detect and characterize adverse events and syndromes.

Individual Exposure Assessment

Epidemiologists generally assess exposure and outcome status related to medical devices by consulting medical records and interviewing patients. Almost all of the challenges to medical device study validity relate to individual exposure assessment. The brand, model, and exact unit used for a particular patient are generally not explicitly written for disposable or durable equipment or short-term implants, but the type of device can generally be inferred from the recorded procedures and knowledge of standard care. In the case of durable equipment, if only one was available at the facility, it is possible to get detailed information on the device. Equipment reused by the same patient may be well recorded in a clinic setting and less well recorded in a home setting. Another problem to be considered with equipment and short-term implants is that device systems (such as for hemodialysis or ventilation) are constructed of many different components, and may include disposable, reused, or durable equipment, a diagnostic device, or

a short-term implant. These components may or may not be the same brand; this is relevant because the performance of a particular component may be affected by the brands of other components. Assessing the critical device for the study hypothesis may require accounting for all the other components in the device system. A piece of durable equipment may present this exposure–assessment dilemma in itself; over time, it may have acquired updated parts from the same or another manufacturer during repair or refurbishment.

For long-term implants, operating room notes and patient records generally have detailed data on the implant. Unfortunately, the advent of bar-coded stickers has resulted in the stickers often being the only written documentation, in the absence of publicly-available codebooks for interpreting the codes. Registries have been formed for various implants, although patients may resist registration if the implant is of a socially sensitive nature. Long-term follow-up, furthermore, can be challenging, especially if the physician following the patient is not the physician inserting the implant.

In the case of diagnostic devices, patient charts generally record the results but not the device used. Depending on the test, home test kit use may go unrecorded.

The newly developed scheme, the Global Medical Device Nomenclature (GMDN), adopted by several countries, is intended to provide standard nomenclature and the basis for a unique identification system for use across countries and manufacturers. However, device codes are not required to be on the individual devices.

When asking a subject about medical device use, recall can be a problem because the patient may not have taken particular note of many of the details. Furthermore, some devices, such as breast implants or impotence devices, are associated with social discomfort, so may be underreported. Interviewers must take special care to encourage full disclosure.

National Population Exposure Assessment

Once a likely relationship between exposure and outcome is discovered, the extent of population exposure must be determined. Public sources of device exposure data include market data, medical care claims, medical records, and population survey data. Market data firms derive their information in various ways, from polling health care providers to collecting device purchase information from a nationally representative set of health care facilities and providers. Market data may express incident or prevalent exposure data. Other examples are described below. Generally, these information sources are much less abundant, reliable, or detailed than comparable types of drug exposure information.

EXAMPLES OF CURRENTLY AVAILABLE SOLUTIONS

Safety Surveillance

Many countries rely on mandatory and voluntary reports for their safety surveillance. In the US, the FDA has taken actions that include recalls, public alerts, required changes to device labeling, and required epidemiologic studies.

The general discussions of reporting systems found elsewhere in this book (Chapters 7 and 8) also apply to adverse *device* event reports. Whether adverse events due to medical devices are more or less likely to be reported than those due to medications is unknown.

Surveys

The population-based general surveys conducted by the US National Center for Health Statistics have provided various data on medical devices (see Table 27.3). *Ad hoc* surveys are also used to study specific devices and adverse effects.

Registries

Registries are sometimes formed to establish cohorts of patients with particular device exposures, generally permanent implants. If a registry is population based, the incidence of device implantation, prevalence of implant

Table 27.3. Some surveys conducted by the US National Center for Health Statistics that include medical device information

Title	General and Device information	Sample size
Medical Device Implant Supplement (MDIS) to the 1988 National Health Interview Survey	"The . . . National Health Interview Survey . . . is based on . . . a continuing nationwide survey by household interview." The survey collects data on demographics and health conditions. The MDIS was administered only in 1988, and collected information on medical implants in all household members.	The sample included 122 000 members of about 47 000 households. The net MDIS response rate was 92%.
1988 National Maternal and Infant Health Survey (NMIHS)	"The NMIHS data file consists of three independent national files of live births, fetal deaths, and infant deaths; and a small supplementary sample of Hispanic live births, fetal deaths, and infant deaths in Texas, and a supplementary sample of live births for urban American Indians. Each mother named on those vital records was mailed a 35-page mother's questionnaire." The data include prenatal and perinatal tests and procedures, such as ultrasound examination. "The 1988 NMIHS can be merged with the 1991 Longitudinal Follow-up." Information about some devices, such as home apnea monitors, is included.	The data file is based on the responses of 10 000 mothers of live births, 3300 mothers of late fetal deaths, and 5300 mothers of infant deaths.
National Health and Nutrition Examination Survey (NHANES)	The data consist of standardized questionnaires, physical examinations, and laboratory testing regarding health and nutrition status and behaviors. Serum latex allergy (IgE) was tested in the 1999–2001 sample of 12–59-year-olds using the AlaStat Microplate method.	10 000 people participated in NHANES 1999–2000 (part of NHANES 1999–2004).
1993 Mortality Followback Survey	"The Mortality Followback Survey Program . . . uses a sample of United States residents who die in a given year to supplement the death certificate with information from the next of kin or another person familiar with the decedent's life history . . . The 1993 survey samples [22 957] individuals aged 15 years or over who died in 1993." The focus areas of the 1993 survey included risk factors for death, disability, unintentional death, and access to and utilization of health care in the last year of life. It provides information on devices used in the home (hospital beds, blood glucose meters, infusion pumps, dialysis machines, etc.).	Almost 23 000 subjects.
2000 National Home and Hospice Care Survey (NHHCS)	"The sampling frame for the 2000 National Home and Hospice Care Survey (NHHCS) consisted of 15 451 agencies classified as agencies providing home health and hospice care." The survey asked about various devices used in home care, including assistive (such as wheelchairs), therapeutic (such as oxygen delivery systems), and diagnostic (such as blood glucose meters) devices.	The sample consisted of 1800 "agencies providing home health and hospice care" services at the time of the survey. Of these, 1425 agreed to participate.

recipients, and adverse event and mortality rates for recipients can also be calculated.

Medical Care Claims

An example of a large US hospital care claims database suitable for medical device epidemiologic study is the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS). Subject to their dependency on procedure codes being adequate indicators of device exposure, studies of the extent of use, adverse events in the same hospitalization, or longitudinal trends can be performed.

Medical Records

Medical record systems vary from entirely paper-based, to partially systematized, to fully automated. Their usefulness for epidemiology is limited by the generally incomplete documentation of adverse medical device events. The assessment of medical device exposure depends on the nature of facility-specific practices regarding the extent of notes or logs.

New Data Collection

New study-specific data collection is also an option that may be crucial to study success. It is often the only option for studies of devices used in the home, because broad surveys may not sample enough device users to be informative and medical records may not capture the information that is required.

Combinations of Techniques

As with other epidemiologic research, combining study techniques can strengthen the results. For example, claims may be used to select a random sample of hospitalizations, patient charts may be abstracted for specific types of adverse events, including some that are related to medical devices, and the claims data may be used to follow the patients after discharge, to note rehospitalization or death within 30 days. Another example of multiple techniques is the use of different survey types as well as complementary study designs (case-control and case-cohort) to address the same questions.

THE FUTURE

In some cases, new device technology is in use before an epidemiologic study of the older technology can be completed. This problem will need to be addressed by faster

epidemiologic techniques, such as quicker assemblage of an exposed cohort after a device is introduced to the market, for observing long-term safety and effectiveness.

Increasing uniformity and automation of medical records will improve the prospects for medical device epidemiology. Records will be most useful when they completely document both device exposure (which should be aided by the unique device identification system under development) and related problems.

Increased recognition of medical devices as an influence on health may allow for major advances in both resources and methodology.

CASE EXAMPLE 27.5: USE OF ALTERNATE STUDY DESIGNS TO ASSESS CORNEAL ULCER RISK ASSOCIATED WITH EXTENDED WEAR OF CONTACT LENSES

Background

- During the 1980s, the ophthalmology and regulatory communities became suspicious that long-term wear of extended-wear soft contact lenses increased the risk of corneal ulcer. The labels suggested wear up to 30 days.

Issues

- Measure the prevalence of use of soft contact lenses.
- Measure the rate of corneal ulcers among users and nonusers of soft contact lenses.
- Among soft contact lens users, identify and measure risk factors for corneal ulcer.

Approach

- Because these lenses are relatively inexpensive, they may be purchased out of pocket from a source outside of the subject's normal health plan. Consequently, full contact lens information is often not available in the subject's main medical records, making exposure assessment difficult to measure from records alone. Thus *ad hoc* studies were required to measure the use and complications from extended-wear soft contact lenses.
- Different survey types as well as complementary study designs (case-control and case-cohort) were used to address the same primary and secondary questions.

(Continued)

The complementary epidemiologic studies were designed together and incorporated both a random sample survey of a population and of all ophthalmologists serving the population, as well as case selection with both hospital and population controls.

- Poggio *et al.* (1989) surveyed the general population to measure exposure to extended-wear soft contact lenses and combined the results with their survey of all ophthalmologists to estimate corneal ulcer rates. Schein *et al.* (1989) used structured interviews of all cases and controls to measure exposure and to assess other details about lens wear.

Results

- Relative risks for corneal ulcer of 5.2 (95% 3.5–7.7) by Poggio *et al.* and 3.9 (95% 2.4–6.5) by Schein *et al.*
- Corneal ulcer risk rises progressively with any lens wear and then each successive night of wear.
- Based on these studies, the FDA found it prudent to have eye care professionals set individual patient maximum wear times, subject to an overall limit of seven days.

Strengths

- The different methods used in each study were probably valid because they resulted in similar effect estimates.
- Case–control and case–cohort studies each have inherent weaknesses; conducting both types strengthened the investigation tremendously.

Limitations

- Having to rely on patient descriptions of their use patterns was a weakness. However, the strong estimated dose–response relationship indicates that this did not significantly impair study interpretation.

Summary Points

- Even when original data collection is required and the outcome of interest is rare, rigorous medical device studies are possible.
- Study design limitations can be compensated for with complementary study designs.
- Results from well-designed epidemiologic studies can result in regulatory changes.

STUDIES OF DRUG-INDUCED BIRTH DEFECTS

Major birth defects, typically defined as those that are life threatening, require major surgery, or present a significant disability, affect approximately 2–4% of liveborn infants. Minor malformations are of lesser clinical importance, and vary considerably in their definition, detection, and prevalence.

Teratogenesis is a unique adverse drug effect, affecting an organism (the fetus) other than the one for whom the drug was intended (the mother). Whatever benefit/risk accrues to the mother, only the fetus is at risk for birth defects.

We are usually completely unaware of a drug's teratogenic potential when it is first marketed. Ignorance about embryologic mechanisms constrains development of predictive *in vitro* tests, and testing in animals is only rarely predictive. Premarketing clinical studies do not provide this information either. Traditionally, women of childbearing age were excluded from clinical studies, specifically because of concerns about potential teratogenicity; newer guidelines encourage their enrolment, but most studies assure that they are at minimal risk of becoming pregnant.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

As noted, the fetus is the “innocent bystander” with respect to its mother's therapy. Further, since roughly half of pregnancies (at least in the US) are unplanned, teratogenic concerns extend to 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.

Drugs Known to be Teratogenic

Human teratogens tend to fall into two broad categories. 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 5–20-fold (e.g., carbamazepine and neural tube defects). The differences between high-risk and moderate-risk teratogens are relevant to how these drugs are considered in the clinical setting.

Various approaches may be applied to the few drugs known to be human teratogens. Such drugs may be prohibited or removed from the general market, as was the case for thalidomide in the 1960s, given its high teratogenic risk without offsetting therapeutic benefits.

Most known teratogens, such as phenytoin and valproic acid, pose moderate risks balanced by clinical benefits. Physicians are expected to discuss the benefits and risks with their patients. Sometimes, a drug may be restricted to prescription by selected physicians.

The third, more recent, approach utilizes a formal risk management program (see also the next section in this chapter) involving education of physicians and patients combined, in some cases, with restricted access to the drug (e.g., thalidomide, isotretinoin).

Drugs for which Teratogenic Risk is Unknown

Most prescription drugs and virtually all non-prescription drugs fall into a much larger category of drugs with unknown teratogenic risk. Labels may offer a general warning against use in pregnancy, but these hardly contribute to rational drug therapy: Where the true teratogenic risk is nil, these warnings deny potentially useful therapy; where the true risk is elevated, the nonspecific warnings offer little practical discouragement to use in pregnancy.

Drugs for which Teratogenesis is Alleged, and Clinical Consequences

At one time or another, many drugs have been alleged to be teratogenic, with profound clinical consequences. For example, about 30 years ago, the widely used anti-nausea drug Bendectin® (Debendox®, Lenotan®) was alleged to cause various birth defects. Despite the lack of support for these allegations, legal concerns led the manufacturer to withdraw the drug from the market. Ironically, the aggregate data for Bendectin have provided the strongest evidence of safety for any medication used in pregnancy. Other effects of unproven allegations include overwhelming guilt among women who give birth to a malformed child and anxiety among currently-pregnant women; consequences of the latter can range from consultations with physicians, to diagnostic procedures (e.g., amniocentesis), to elective termination of the pregnancy.

The Fallacy of "Class Action" Teratogenesis

While structure/activity relationships shared by members of a given drug class can help to predict a given class member's

efficacy and adversity, this cannot be assumed for teratogenesis because we cannot know whether the responsible component is the "class" or that part of a drug's structure that differentiates one class member from another. Thus, we cannot project the teratogenicity (or safety) of one class member onto others.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Sample Size Considerations

"Birth defects" is not a single, homogeneous outcome, and teratogens do not uniformly increase the rates of all birth defects, but rather increase the rates of selected defects. Defects vary widely in terms of gestational timing, embryologic tissue of origin, and mechanism of development. Thus, one would predict that 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 teratogenic mechanism.

The need to focus on *specific* birth defects dramatically affects sample size requirements. For example, a cohort study of a few hundred exposed pregnancies might be sufficient to identify a doubling of the overall rate of "birth defects"; ruling out a doubling of the overall rate would require larger numbers, but is still within the same order of magnitude. However, any specific defect is rare, affecting 1 per 1000 live births (e.g., oral clefts) to 1 per 10000 or fewer (e.g., biliary atresia). For a cohort study to *detect* a doubling of risk for a relatively common specific birth defect (e.g., 1/1000) requires over 20000 exposed pregnancies. To *rule out* a doubling of risk for the same defect, one would need a far larger sample size of exposed pregnancies.

Exposure

Non-Prescribed Drugs

Over-the-counter (OTC) drugs present a unique situation. By definition, use of OTC drugs does not require physician involvement. Like consumers and physicians, women of childbearing age view OTC drugs as safer than prescription products, and they may assume that the same is true for use of these drugs in pregnancy. Though prescription drugs generally become OTC based on a history of wide use and safety, this history rarely includes systematic information on potential human teratogenicity. This is particularly true for drugs that became available OTC decades ago.

Although the teratogenic effects of OTC drugs are largely unstudied, these agents have been used widely in pregnancy

for decades, and the more recent increase in herbal product use has raised additional concerns. These exposures should be a focus of pharmacoepidemiologic research.

Recall Bias

Deriving exposure information from maternal interviews raises concern about recall accuracy and potential bias (see Chapter 15). Because of feelings of guilt, the mother of a malformed infant may be more likely than the mother of a healthy infant to accurately recall her pregnancy exposures, and the difference in recall accuracy could lead to false positive associations. However, the simple possibility of recall bias does not invalidate interview-based studies, and there are various approaches to minimizing this problem.

Restricting analysis to mothers of malformed infants (whether they are cases or controls) limits the likelihood of false positive results due to biased recall among mothers of malformed infants. The likelihood of a false negative can be minimized by including a wide range of malformations among controls, since it is unlikely that a drug increases the risk of most specific defects.

Open-ended questions invite differential recall between mothers of malformed and normal infants, whereas focused questions are more likely to yield complete information. Recall is also substantially increased when women are asked about use according to various indications and when drugs are asked by specific names. Still, 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.

Outcome

Though birth defects are often classified by organ system (e.g., “musculoskeletal”), whenever possible they should be classified on the basis of the embryologic origin for a given defect. For example, neural crest cells form various structures of the face/ears, heart, and neural tube, and the retinoid isotretinoin, which interferes with neural crest cell migration/development, produces malformations of the ear, heart, and neural tube.

Confounding

Potential confounders that are routinely considered in teratologic studies include maternal age, race, geography, and socioeconomic status. However, additional potential confounders must be identified based on understanding the epidemiology of specific defects. Also, since medication use may be associated with various health behaviors (e.g.,

vitamin use is more common among nonsmokers), one may need to consider various health behaviors, including smoking, alcohol, and nutrition, in studies of certain exposures and outcomes. Further, it is critically important to separate risks due to the drug from risks associated with the condition for which the drug is taken, known as “confounding by indication” (see Chapter 21).

Finally, there is the important possibility of pregnancy termination. As 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, and the likelihood of termination may be related both to the outcome under study and the use of a given drug.

Biologic Plausibility

How does one evaluate the importance of biologic plausibility for newly observed associations? While a requirement that every association have an identifiable biologic mechanism would have led to dismissal of most accepted human teratogens, some aspects of biologic plausibility must be met. For example, it is implausible that a defect could be caused by an exposure that first occurs after the gestational development of the defect. It is also unlikely that an exposure would produce defects that span gestational timing from preconception to late pregnancy and do not share embryologic tissue of origin. Thus, though we cannot dismiss hypotheses simply because they lack a biologically plausible explanation, until they are supported by subsequent studies such hypotheses must be considered more speculative than hypotheses with strong biologic plausibility.

EXAMPLES OF CURRENTLY AVAILABLE SOLUTIONS

Cohort and case-control designs are the favored approaches used to generate and test hypotheses regarding drugs and birth defects.

Cohorts

Three types of cohorts relevant to the pharmacoepidemiologic study of birth defects include studies designed to follow large populations exposed to various agents, 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 following a population of pregnant women with collection of data on exposures, outcomes, and covariates. 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. The strength of this approach is 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. However, a major weakness of even a cohort this large is the relatively few infants with *specific* malformations. While the CPP included approximately 2200 infants with “major malformations,” there were only 31 with cleft palate (CP) and 11 with tracheoesophageal fistula (TEF). This weakness is compounded by limited numbers of women exposed to most drugs. For a drug taken by as many as 10% of the women, the expected number of exposed infants with CP and TEF would be 3 and 1, respectively; if a drug were used by 3%, the expected intercepts would be 1 and 0.3. Such a cohort may be large enough to identify some high-risk teratogens, but power is usually inadequate 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, the inordinate costs of such intensive efforts limit enrollment and data collection to a few years. Because of changing patterns of drug use over time, the clinical relevance of the available data diminishes.

Use of Data Sets Created for Other Purposes

Increasing attention has focused on cohorts identified from administrative databases produced for purposes other than epidemiologic research by organizations or governments involved in medical care (see Chapters 11 and 12). 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 include large populations, and a few may have good reporting of malformations. Despite their overall size, these databases may include few subjects with specific malformations who were exposed to a particular drug *in utero*. Further, information is usually absent on important confounding variables (e.g., smoking, alcohol, diet, OTC drugs), and the quality of diagnosis data is often quite limited. While administrative datasets may have value in identifying high-risk teratogens, their value for moderate-risk teratogens is modest.

Follow-up of Selected Exposures

Cohorts of women exposed to specific drugs can be developed by enrolling pregnant women in pregnancy registries, either by physicians or by the women themselves. Registries, whether operated by a manufacturer or a teratogen information service, can identify a woman exposed to a drug of interest early in pregnancy and, most importantly, identify and enroll the woman before the pregnancy outcome is known. Registries that are in direct contact with pregnant women can also prospectively collect other information, such as data relating to other exposures and potential confounding variables.

Cohorts of a few dozen to a few hundred exposed pregnancies are highly efficient and effective for identifying—and ruling out—high-risk teratogens. On the other hand, such cohorts are quite limited in their ability to identify a drug as a moderate-risk teratogen or to rule out such an effect.

Registries may be limited by problems of self-referral bias and losses to follow-up. In addition, comparators may be absent or inappropriate, and the registry data may not allow exploration of confounding by indication.

It is difficult for any cohort study, though, to achieve sample sizes sufficiently large to provide critical information not simply on the risk of birth defects overall, but rather on risks of specific defects, most of which have background rates ranging from 1 per 1000 to 1 per 10 000.

Case-Control Studies

The rarity of birth defects overall, and particularly specific defects, argues for the use of the case-control design when exposure prevalence is sufficiently high. Such studies may be conducted on an *ad hoc* basis or within the context of case-control surveillance (see Chapter 9). Examples of the latter include the longstanding “Birth Defects Study” and the more recently established National Birth Defects Prevention Study, involving several state birth defects surveillance programs and coordinated by the CDC. From the perspective of specific birth defects, case-control studies can have the statistical power required for the assessment of both risk and safety. By obtaining information directly from the mother, they also can capture information on critical covariates, such as smoking, alcohol, diet, and use of non-prescription drugs.

Statistical power is a major strength of the case-control approach, but power does not assure validity. A concern with this approach is recall bias; however, the simple *possibility* of such bias does not invalidate a case-control study; rather, its potential requires careful efforts to consider and minimize such bias in the study population.

An Integrated Approach

Although not widely appreciated, the need to identify high-risk and moderate-risk teratogens can be efficiently achieved by combining the approaches described above. Cohorts of exposed subjects, preferably in the form of a centralized pregnancy registry, can identify the few high-risk teratogens in a timely way. The extremely large risks of drugs such as thalidomide or isotretinoin tend to overwhelm the methodologic limitations inherent in these approaches. 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, and the same design can provide relatively stable estimates of a drug's safety.

THE FUTURE

Integration of Epidemiology and Biology

Advances in molecular biology will markedly enhance our ability to classify defects in biologically meaningful categories and facilitate the development of biologically plausible hypotheses. Even more promising is the rapid expansion of research on genetic polymorphisms of drug-metabolizing enzymes (DMEs). Human teratogens do not produce malformations in all (or even most) exposed fetuses, and this "incomplete penetrance" is likely due to differences in host susceptibilities, such as the host's handling of drugs. Polymorphic DMEs are being identified at a rapid pace, and investigators are increasingly adding data banks of buccal cells or blood samples to ongoing studies of risk factors for birth defects. Understanding genetic polymorphisms will permit identification both of population subsets at increased risk for certain birth defects and of drugs that warrant particular study. In advance of pregnancy, women may thus be screened for genetic polymorphisms that 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 of becoming pregnant.

The Legal and Regulatory Climate

Pharmacoepidemiologic studies of teratogenesis require access to information that identifies women who have become pregnant, their medication exposures, and details of the outcomes of those pregnancies. Disclosure of such information to researchers has become highly contentious

in many countries. For case-control studies in particular, the enrollment of malformed and/or normal subjects requires that identifying information be made available to researchers, who then contact eligible subjects to invite them to participate in an interview.

Though there is little evidence that medical researchers in this area have compromised confidentiality, epidemiologic studies may be compromised or eliminated by privacy/confidentiality regulations (see Chapter 19). It is critical that researchers engage their communities about the public health value of epidemiologic research and the need for balancing privacy concerns with the need to provide critical information on the risks and safety of medications taken in pregnancy.

Hope for an Integrated Approach

Future studies of birth defects will undoubtedly focus increased attention on issues of statistical power, validity, and secular changes in use not only related to prescription drugs, but also to OTC and herbal products. Although the thalidomide debacle did much to stimulate research and regulatory attention on the adverse effects of medications, it is ironic that in the more than 40 years since that teratogenic disaster, drugs have yet to come under systematic study for their potential teratogenic risks. Consequently, information available to pregnant women and their health care providers about the safety of drugs in pregnancy is not much better today than it was 50 years ago. This situation need not persist—systematic approaches to pharmacoepidemiologic research in this important area can be established simply by coordinating proven methodologies. In the US, for example, our knowledge of each drug's risk and safety to the fetus can be dramatically improved, cost-efficiently, by developing a comprehensive surveillance system that combines the complementary strengths of already established cohort and case-control surveillance infrastructures.

CASE EXAMPLE 27.6: TESTING HYPOTHESIZED TERATOGENIC EFFECTS

Background

- In the late 1970s, the anti-nausea drug Bendectin® (Debendox®; Lenotan®: doxylamine, dicyclomine, and pyridoxine) was widely used to treat nausea and vomiting in pregnancy.

- Legal claims based on allegations of the drug's teratogenicity ultimately resulted in the manufacturer removing it from the market.

Issue

- Concern about the possible teratogenic effects of the drug were raised by studies that suggested increased risks of selected cardiac defects and oral clefts among the babies of mothers who had taken the drug in the first trimester.
- In both studies, exposure among mothers of cases was compared to exposure among mothers of normal infants, and there were questions about the rigor and symmetry regarding collection of exposure information.

Approach

- Utilizing data from an ongoing case-control surveillance program of specific birth defects in relation to medication use in pregnancy, researchers identified cases with selected cardiac defects and cases with two kinds of oral clefts and compared maternal Bendectin exposure among those cases to that among mothers of controls with various malformations other than those in the cases.

Results

- Among the 970 malformed controls, the prevalence of first trimester exposure was 21%.
- Case groups ranged in size from 98 to 221 infants.
- Risk estimates ranged from 0.6 to 1.0.
- All upper 95% confidence bounds were 1.6 or less.

Strengths

- The existence of the case-control surveillance database provided sufficient power (for this common exposure) to test the hypotheses without the need for further data collection.
- Sample sizes were large enough to provide tight confidence bounds.
- Direct interviews with mothers of study subjects provided information on important potential confounding variables.
- Use of malformed subjects as controls minimized the risk of biased recall.

- Prevalence of exposure among controls was similar to that identified by sales data.

Limitations

- Malformed controls could include defects caused by the drug of interest.
- The possibility of very small risks could not be excluded.
- The possibility of residual confounding remains.

Summary Points

- Recall bias can affect findings of case-control studies, particularly in comparisons between malformed and normal infants when exposure information is not rigorously and symmetrically collected.
- Ongoing case-control surveillance approaches, by accruing large numbers of subjects with specific malformations with detailed exposure and covariate data, can efficiently test hypotheses without the need to mount new studies that would take years to complete.
- The impact of false positive teratogenic studies can be considerable.

PHARMACOEPIDEMIOLOGY AND RISK MANAGEMENT

The use of the term “risk management” (RM) in the setting of prescription drug safety is a relatively recent development and a consensus definition for it has not yet emerged. Risk management represents a growing awareness of the need for continuing evaluation of risks related to drug products and the implementation of plans to minimize those risks after marketing. The concept has emerged and gradually developed, especially in response to the challenges created by accelerated review and approval times for marketed drugs, as well as drug product withdrawals for safety reasons in the US. Often, RM represents those interventions that must be in place and must be effective in order to shift the balance of benefit to risk from an unfavorable and unacceptable position to one that is favorable and acceptable. A corollary of this formulation is that, in these situations, in the absence of a successful RM program, the balance of benefit to risk for a drug requiring an RM program would be unfavorable and unacceptable. Regardless, the ultimate goal of RM is to enhance the safe use of medicines by optimizing the balance

of benefit and risk. This can be accomplished by increasing the benefit and/or decreasing the risk associated with use of a drug product. Alternatively, it can be accomplished by limiting the use of the drug to those most likely to need it.

Numerous interventions have been employed to attempt to improve the balance of benefit to risk of specific marketed drug products. Examples of some specific interventions have been discussed in the preceding sections on Physician Prescribing and on Drug Utilization Review. The most common methods used to attempt to improve the balance of benefit to risk of specific drug products include professional labeling and package inserts, special letters sent to health care professionals, specific educational and training programs, and patient-package inserts and medication guides. Other strategies include the signing of informed consent forms by patients prior to drug therapy, registration of health care practitioners before they are “certified” to prescribe or dispense a given drug, requiring the performance of certain laboratory tests prior to prescribing, and establishing registries of physicians, pharmacists, or patients who, respectively, prescribe, dispense, or use a given drug product. Access to the drug in question may be restricted. The use of special packaging, imposing limits on prescription size, and reformulation of drug products have also been used.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Each RM program should have specific, clinically meaningful, and measurable goals. These goals should be directly relevant to the safety problem and should reflect the desired *health outcomes* of the RM program, rather than just process measures or surrogates for the health outcomes of interest. Because there can be great uncertainty regarding the effectiveness of RM programs, their use can be conceptualized as experimental in nature.

Risk Identification and Characterization

Randomized controlled trials (RCTs) have a number of limitations, including relatively small size, short duration, and recruitment of highly selected and hence unrepresentative patients. Consequently, well-designed epidemiologic studies frequently offer the only way to characterize the safety risks associated with a given drug.

Postmarketing

In the early stages of marketing of a new drug, observational studies using computerized databases are often not practicable because there has not been sufficient patient exposure

to the drug of interest. From an RM perspective, this means that during the early postmarketing phase most information about risk will be obtained from spontaneous or published case reports.

Product Usage

From the start of marketing, drug use can be monitored. Drug use in inappropriate age groups, at higher doses, for longer lengths of time than labeled, or for off-label use might signal the emergence of a potential safety problem requiring RM.

Case Reports

Review of well-documented case reports can provide useful insight into the spectrum, severity, and natural history of an adverse drug effect, as well as its reversibility, predictability, and preventability (see Chapters 7, 8, and 17). Understanding these factors should be a critical component of any decision to design or implement an RM program.

Phase IV and *ad hoc* Postmarketing Epidemiologic Studies

Pharmacoepidemiologic studies are sometimes relied upon to characterize further the nature and magnitude of safety signals arising during pre-approval clinical trials or from postmarketing spontaneous case reports. Limitations in the methods and design of these studies can undermine their power and utility, frequently leading to negative findings.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Properly Utilized, Pharmacoepidemiology Risk Management Goal Setting

A comprehensive understanding of factors contributing to drug safety risk is critical to establishing RM goals. High quality case reports may suggest junctures along the causal path where intervention might be useful and effective. Careful review of drug usage data may also contribute by documenting patterns of use that are intrinsically unsafe or that are off-label.

The primary goals of a RM program must be explicit, relevant, and measurable. An explicit goal is unambiguous and can be subjected to evaluation regarding success or failure of goal attainment. A relevant goal usually will be a direct health outcome immediately reflective of the *event* that the RM program is seeking to eliminate or minimize. Process goals that assess incremental aspects or components of a

RM program would not be relevant. Finally, a measurable goal is a health outcome that can be measured, permitting conclusions to be made regarding goal achievement.

Risk management programs are especially warranted for the severe or serious end of the adverse effect spectrum, typically those that are potentially life threatening, or result in hospitalization, disability, or death. In these settings, the requirement that an RM program be implemented is an admission that, without effective RM, the balance of benefit to risk would be unacceptable.

Risk Management Design and Evaluation

Risk management programs often represent interventions of uncertain and unknown effectiveness that will be applied to large numbers of patients within the “laboratory” of the postmarketing health care setting. If the design of an RM program does not foster the collection of accurate and reliable data on the health outcome of interest, there is no way to determine if the risk has been adequately managed and the balance of benefit to risk optimized.

Examples of Currently Available Solutions

While RM may be considered to cover a broad range of safety issues and concerns, the above discussion implicitly addresses RM programs for the severe or serious end of the adverse effect spectrum, typically those that are potentially life threatening, or result in hospitalization, disability, or death. In these settings, the requirement that an RM program be implemented is a tacit admission that, without effective risk management, the balance of benefit to risk would be unacceptable. In this setting, implementation of an effective RM program is necessary to establish an acceptable balance. Examples of drug risks and efforts to manage them by minimizing harm are provided in Table 27.4.

Effectiveness of Risk Management Efforts

Changes to product labeling and the posting of letters to health care providers about product risk, while absolutely necessary, do not alone appear to result in successful risk management, that is, optimized patient safety via changing physician or patient behaviors.

Risk management programs supported by national legislation or a nationally coordinated comprehensive campaign, and those relying on “engineering-based” approaches, have shown the greatest success at improving major health outcomes in the face of serious drug safety risks. “Engineering-based” refers to interventions that do not rely purely on patient or physician compliance, cooperation, or

knowledge. Examples include systems that regulate access to the drug product or that impose limits on access to dangerous quantities of the drug product. Manufacturing processes that physically alter or change the drug product to reduce or eliminate its intrinsic toxicity are yet another approach.

THE FUTURE

Pharmacoepidemiology should be an intrinsic component of the RM process, from identifying and characterizing product risk to evaluating the success or failure of a product’s RM program. Currently, the absence of reliable evidence of RM success usually results in a default position favoring the *status quo*. This absence of evidence allows ineffective RM programs to continue while pharmaceutical companies and drug regulatory authorities mistakenly assume that they have successfully mitigated a product’s risk.

If a safety risk is serious enough to mandate a special RM program, several principles become operational. Borrowing from the world of clinical trials, exposing large numbers of patients to an apparently ineffective drug within a Phase III clinical trial would not be justified. It is also difficult to imagine an institutional review board approving the use of an untested or ineffective therapy in a clinical trial involving human subjects. In the RM setting, analogous principles should apply. When a serious postmarketing safety issue arises, involving tens if not hundreds of thousands of treated patients each year, the imperative to protect patients from harm becomes more urgent. Reliance on untested or unproven RM methods to reduce risk, or optimize the balance of benefit to risk, could be considered inappropriate or even unethical.

Another principle governing clinical trials is that appropriate data be collected to assess the success of a treatment. To not do so would be scientifically inappropriate and unacceptable. Yet, to date, most enacted RM programs have failed to collect the appropriate data or even insist that it be collected.

The potential pitfalls of (i) relying on RM interventions that have little or no probability of success or (ii) not collecting the right types of data to document optimization of the balance of benefit to risk can all too easily threaten the value and integrity of the practice of RM. The effectiveness of RM programs cannot be assumed, but must be measured and documented. Science-based efforts in this field must ensure that RM not become a strategy to enable the continued market availability of dangerous drugs, without addressing the underlying safety problem with an effective intervention.

Table 27.4. Summary of risk management programs and their effectiveness as qualitatively assessed by the authors

Drug	Safety concern	Risk management interventions	Effect of interventions
Acetaminophen	Overdose, acute liver failure, death	National law in UK mandating blister packs and sublethal package size	↓ hospitalizations, mortality, liver transplants
All drugs	Pediatric overdose, death	National law mandating child-resistant packaging	↓ mortality from accidental poisoning
Aspirin	Reye's syndrome	Prolonged multifaceted national campaign at Federal government level	↓ in incidence, mortality
Bromfenac	Acute liver failure	Boxed warning in label, Dear HCP letter	No change in prescribing behaviors; continued reporting of liver failure
Cisapride	Fatal cardiac arrhythmias	Multiple label changes and warnings, multiple Dear HCP letters, multiple articles in medical literature	± small changes in prescribing behavior; continued reports of TdP
Clozapine	Agranulocytosis	Restricted access with mandatory linkage of normal lab result and registered prescription release	↓ incidence of agranulocytosis
Isotretinoin	Pregnancy exposure	Multiple label changes and warnings, multiple Dear HCP letters, multiple public advisory meetings, multiple articles in medical literature, multiple intensive educational campaigns, signed informed consent, contraindicated use, qualification stickers, physician registry, pregnancy testing	No evidence of reduced pregnancy exposures; poor compliance with pregnancy testing; >2-fold increase in use of drug in women of child-bearing age (off-label use)
Pemoline	Acute liver failure	Label changes, multiple Dear HCP letters, second-line therapy, liver enzyme monitoring	Persisting high level of 1st-line use; poor compliance with liver enzyme monitoring; no reliable data on liver failure occurrence
Terfenadine	Fatal cardiac arrhythmias	Multiple label changes, multiple Dear HCP letters, replacement by non-cardiotoxic metabolite, fexofenadine	↓ co-prescribing with contraindicated macrolides and azole anti-fungals; persisting 1–3% same-day co-prescribing; continued reporting of cases of TdP
Troglitazone	Acute liver failure	Multiple label changes, multiple Dear HCP letters, public advisory meeting, national media attention, liver enzyme monitoring. Removal from market in 2000	Poor compliance with liver enzyme monitoring; evidence that monitoring would not be effective had it been performed; continued reporting of liver failure

CASE EXAMPLE 27.7: **EFFECTIVENESS OF RISK MANAGEMENT** **WITH TROGLITAZONE**

Background

- Within 6 months of approval, troglitazone, an oral hypoglycemic agent, was found to increase the risk of acute liver failure. As a risk management strategy, the FDA implemented a series of labeling changes

recommending that physicians perform baseline and monthly serum liver enzyme testing.

Question

- Did the FDA's risk management program prevent the occurrence of acute liver failure?

Approach

- Systematic review of adverse event case reports.

- Population-based cohort study of hospitalization for acute liver injury.
- Population-based cohort study of serum liver enzyme monitoring in patients treated with troglitazone.

Results

- Troglitazone substantially increased the risk of acute liver failure and hospitalization for acute liver injury.
- The FDA risk management efforts did not result in sustained or meaningful improvement in performance of liver enzyme monitoring by physicians.
- Even if monthly liver enzyme monitoring had been performed, it is unlikely that it would have prevented many, or perhaps any, cases of acute liver failure.

Strength

- Comprehensive evaluation of liver failure risk and the effect of repeated waves of risk management interventions with troglitazone.

Limitations

- Incomplete information from adverse event case reports.
- Reliance on claims data could have resulted in slight underestimation of liver enzyme monitoring performance.

Summary Points

- Risk management interventions, including repeated labeling changes, multiple warning letters to physicians, and other educational efforts, did not result in meaningful or sustained improvement in liver enzyme monitoring or in the safety profile of troglitazone.
- Based on analysis of case reports, it is unlikely that monthly liver enzyme monitoring would have prevented the occurrence of acute liver failure with troglitazone.
- The effectiveness of periodic liver enzyme monitoring to prevent severe liver injury with other drugs was not addressed here.
- Risk management interventions should be pilot-tested before widespread implementation and should be evaluated for effectiveness.

THE USE OF PHARMACOEPIDEMOLOGY TO STUDY MEDICATION ERRORS

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. Given the prevalence of prescription medication use, it is not surprising that preventable adverse drug events are one of the most frequent types of preventable iatrogenic injuries. The Institute of Medicine report, *To Err Is Human*, suggested at least 44 000–98 000 deaths nationally from iatrogenic injury. If accurate, this would mean that there are about 8000 deaths yearly from adverse drug events and 1 million injuries from drug use.

Safety Theory

One theory, borrowed from psychology and human engineering, focuses on the man–machine interface, where providers must make decisions using data from multiple monitoring systems while under stress and bombarded by interruptions.

Industrial quality theory suggests that most accidents occur because of problems with the production process itself, not the individuals operating it. This theory suggests that blame for problems can be misplaced, and that although “human errors” occur commonly the true cause of accidents is often the underlying systems that allow an operator error to result in an accident. *Root cause* analysis can be used to define the cause of the defect. While sometimes an individual who makes repeated errors is to blame, most errors resulting in harm are made by workers whose overall work is good. To make the hospital a safer place, a key initial step is to eliminate the culture of blame, and build a culture of safety. Errors and adverse outcomes should be treated as opportunities for improving the process of care through system changes, rather than a signal to begin disciplinary proceedings.

Systems changes for reducing errors can greatly reduce the likelihood of error and probably, in turn, of adverse outcomes. Within medicine, much of the research has come from anesthesia, which has made major improvements in safety. Examples of successful systems changes in medication delivery include implementation of unit dosing, computerized physician order entry, barcoding of medications, and implementation of “smart pumps” that can recognize what medication is being delivered. These technologies can track medication use and, more importantly, the frequencies and types of warnings as they go off.

Overall, the area of safety has a different philosophy and several different tools than classic epidemiology. For improving safety, culture is extremely important, and tools such as root cause analysis and failure mode and effects analysis—which can be used to project what the problems with a process may be before they occur—are highly valuable. When combined with epidemiological data, such tools may be extremely powerful for improving the safety of care.

Patient Safety Concepts, as Applied to Pharmacoepidemiology

While pharmacoepidemiologic techniques have most often been used to study the risks and benefits of drugs, they can also be used to study medication errors and preventable adverse drug events (i.e., those due to errors). *Medication errors* are defined as any mistake in the medication use process, including prescribing, transcribing, dispensing, administering, and monitoring. *Adverse drug events* (ADEs) constitute harm resulting from medication use, and may be preventable or not. By definition, preventable ADEs are associated with an error, while nonpreventable ADEs are not. An example of a preventable ADE is a patient who is prescribed an antibiotic despite a known allergy and develops a rash, whereas a nonpreventable ADE would be a patient with no known drug allergies who is prescribed an antibiotic and develops a rash. Finally, *near misses* or *potential ADEs* are medication errors with high potential for causing harm but do not, either because they were intercepted before reaching a patient or because the error reached the patient who fortuitously did not have any observable untoward sequelae. An example of the former is a prescription written for an overdose of a narcotic that is intercepted and corrected by a pharmacist before medication dispensing. An example of a non-intercepted near miss is a patient administered a two-fold overdose of a narcotic but without any consequence, such as respiratory depression or sedation.

Many of the early medication error and ADE studies were performed in the hospital setting. In the inpatient adult setting, patients are vulnerable to medication errors because of their medical acuity, the complexity of their disease process, and medication regimens, as well as at times because of their age (e.g., the elderly are particularly susceptible). In pediatric drug use, the system-based factors that may contribute to a higher rate of near misses include the need for weight-based dosing and dilution of stock medicines, as well as decreased communication abilities of young children.

Increasing knowledge is being gained about errors in the ambulatory setting, although research lags behind the inpatient setting due to the difficulties of accessing patients once they leave a doctor's office. Finally, some work is emerging about errors at the point of transition from hospitals to ambulatory settings. Handoffs, in which the clinical care of a patient is transferred from one health provider or entity to another, are always vulnerable to errors. Most studies in this area rely on primary data collection, including prescription and chart review, or direct observation by study staff of clinical care interactions primarily in the inpatient setting. Such data collection is very time and labor intensive. Typically it can only be successfully undertaken in the setting of research studies where large amounts of resources are available for data collection, including the training of data collectors to ensure inter-rater reliability. Comparisons among studies are challenging because of variations in data quality and methodology.

Recently, other types of pharmacoepidemiologic techniques, such as claims-based evaluations, have been used. Claims-based evaluations allow the ascertainment of errors in multiple parts of the medication use process, including ordering, transcribing, dispensing, and, in the outpatient setting, compliance. However, it is difficult to hone in on the actual stage at which an error occurred using solely claims-based data. For example, if a claims-based evaluation demonstrates that a patient was given an overdose of a medication, it is difficult to determine if this occurred due to a drug ordering, dispensing, or transcribing error.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Medication errors can occur at any stage of the medication use process, including prescribing, transcribing, dispensing, administering, and monitoring. Of these stages, prescribing errors in the hospital have been documented to cause the most harm, although errors at any stage can do so, and monitoring errors (i.e., errors caused by lack of proper monitoring) are quite prominent outside the hospital. The greater proportion of harmful errors at the drug prescribing stage may be a consequence of the data collection methodology employed in these studies, which were multi-pronged but excluded direct observation, the most sensitive technique for administration error detection.

Important types of errors include dosing, route, frequency, drug-allergy, drug-drug interaction, drug-laboratory (including renal dosing), drug-patient characteristic, and drug administration during pregnancy. Although

these errors occur most frequently at the drug ordering stage, they can occur at any stage in the medication use process.

In several studies, *dosing errors* have represented the most frequent category. To determine whether or not a dosing error is present, most often some clinical context is needed, for example the patient's age, gender, weight, level of renal function, prior response to the medication (if it has been used previously), response to other similar medications, clinical condition, and often the indication for the therapy. While many of these data elements can be obtained from review of the medical chart, many are not typically available from claims data alone.

Route of administration problems also represent a common type of error. Many drugs can be given by one or a few routes and not by many others. Some such errors—such as giving benzathine penicillin that contains suspended solids intravenously instead of intramuscularly—would often be fatal, and though they have caused fatalities they are fortunately very rare. Other route errors—such as grinding up a sustained released preparation to give it via a jejunostomy tube—are much more frequent, and can have very serious consequences. Route errors are especially problematic at the administration stage of the medication use process, and administration errors are both difficult to detect and much less often intercepted than prescribing errors. The best approach for detecting administration errors has long been direct observation.

Dosing frequency errors can occur either at the prescribing, dispensing, or administration stage. While these errors probably cause less harm cumulatively than dose or route errors, they can be problematic. Some frequency errors at the prescribing or dispensing stage can be detected even with claims or prescription data. Such errors have greater potential for harm when drugs are given with a greater frequency than intended. However, the therapeutic benefit may not be realized when given with too low frequency, and extremely negative effects can occur for some drugs, for example with antiretrovirals, to which resistance develops if they are given at a low frequency.

Allergy errors represent a particularly serious type of error, even though most of the time when a drug is given to a patient with a known allergy the patient does well. Allergy errors typically cannot be detected with claims data, since allergy information on patients is not available. Thus, these errors have to be detected either through chart review, which is laborious, or more often through electronic record data.

Drug-drug interaction exposures represent an interesting and difficult area, both for research and interventions,

to decrease errors. While many interactions have been reported, the severity varies substantially from minor to life-threatening. If a conscious decision is made to give patients two medications despite the knowledge that they interact, this cannot be considered an error except in very limited circumstances, for example with meperidine and monoamine oxidase inhibitors. Also, it is legitimate to give many medications together despite clear interactions with important consequences if there are no good alternatives, or if dose alterations are made, or if additional monitoring is carried out (e.g., with warfarin and many antibiotics). However, the necessary alterations in dosing or additional monitoring are often omitted, which can have severe consequences. It is possible in large claims data sets to detect situations in which simultaneous exposures appear to have occurred, but not possible to determine if this actually occurred, as a physician may give patients instructions to cease the use of one of the drugs.

Drug-laboratory errors represent an important category of errors, but can be difficult to detect electronically because of poor interfaces between laboratory and pharmacy information. Such errors are relatively straightforward to identify when large pharmacy and laboratory databases can be linked, although again assessment of clinical outcomes is difficult unless these data are also available.

Renal dosing errors represent a specific subtype of drug-laboratory errors and are especially important. In one large inpatient study, nearly 40% of inpatients had at least mild renal insufficiency, and there are many medications that require dosing adjustment in the presence of decreased glomerular filtration. In that study, without clinical decision support, patients received the appropriate dose and frequency of medication only 30% of the time.

Many studies of drug-patient characteristic checking have focused on the use of medications in the presence of specific diseases. However, in the future, genomic testing will undoubtedly dominate, as many genes have profound effects on drug metabolism. Currently, few large data sets can be linked with genotype information, but this is becoming increasingly frequent in clinical trials and a number of cohorts are being established as well.

Another important type of error results inadvertently from *system-based interventions* such as the introduction of information technology: for example, automated pharmacy systems, featuring computer-controlled devices that package, dispense, distribute, and/or control medications; these also have the potential to reduce administration errors. Although these systems generally reduce errors, one study demonstrated an increase in errors with a device that allowed nurses to obtain any medicine stored for any patient and

did not integrate the computerized medication profiles of patients.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Information Bias

In performing drug analyses, the present conventions preclude the determination of total daily dose in several ways. Physicians may prescribe a greater amount of medicine than is required for the time period prescribed. For example, if a patient requires 50 mg of atenolol per day, the doctor may actually write a prescription for 100 mg of atenolol per day and verbally convey instructions to the patient to divide the pills. This is particularly problematic with drugs that must be titrated to the appropriate therapeutic dose. If either physicians or pharmacists were required to document an accurate total daily dose, this would improve the ability to perform research.

Another important methodological issue is measurement of patient adherence to medications. Since prescribing and dispensing data are seldom jointly available, determining patient adherence is extremely difficult. Improving clinician access to data from pharmacy benefit managers might be very useful, as might the availability of electronic prescription data to pharmacies.

Many medications are contraindicated in pregnancy. Here, the greatest difficulty for the investigator is assessing whether or not the patient is actually pregnant at the time of the exposure, although this can be assessed retrospectively by identifying the date of birth, assuming a term pregnancy, and then working backward. The outcomes of interest are often not represented in ways that make it easy to perform analyses, although data on medication exposures and on births are readily available and can often be linked.

Another important piece of clinical information for pediatrics is a child's weight. Most pediatric medications are dosed on the basis of weight. Standardized documentation of this information is unavailable, hindering not only analyses of pediatric dosing but also actual dosing by pediatricians.

A final issue is the coding of allergies. It is important for both clinical care and research that allergies are differentiated from sensitivities or intolerances through codes rather than free text. Continued drug use in the presence of drug sensitivity may be perfectly appropriate, whereas the same treatment in the presence of an allergy is likely to be an

error. It is particularly important that severe reactions, such as anaphylaxis, are clearly coded and identifiable in the medical records. New allergies need to be captured in better ways. The eventual aim is to have one universal allergy list in an electronic format for each patient, rather than multiple disparate lists.

Sample Size Issues

Sample sizes are often small in medication error and ADE studies, primarily due to the high costs of primary data collecting studies. Electronic databases will be an important tool to improve sample sizes in a cost-effective manner. Computerized physician order entry systems, electronic health records, test result viewing systems, computerized pharmacy systems, barcoding systems, pharmacy benefit managers, and claims systems will all be important sources of such data. There will be important regulatory issues that will need to be addressed before actual construction and use of these systems.

Generalizability

Many existing medication error studies have limited generalizability due to setting or methodology. For example, many studies have been performed in tertiary care, academic hospital settings. It is unclear how findings from this setting translate to other settings. Also, methodologies vary widely from study to study, hindering comparisons.

EXAMPLES OF CURRENTLY AVAILABLE SOLUTIONS

Several data sources can be used to assess the frequency of medication errors: claims data; claims data linked with other types of clinical information such as laboratory data; electronic medical record information including that from computerized order entry or pharmacy systems; chart review; and direct observation. Spontaneous reporting can also be used, but underreporting is so great that it is only useful for getting samples of errors, and cannot be used to assess the underlying rate of medication errors in a population.

Claims data have the great advantage that they can be obtained for very large numbers of individuals. However, it cannot be determined with certainty whether or not the patient actually consumed the medication, and clinical detail is often minimal, making it hard to ask questions that relate to a patient's clinical condition. Searches can be performed for specific diagnoses, but the accuracy of coding is limited

for many of the diagnoses of interest, for example renal failure and depression, as well as for outcomes such as renal insufficiency.

Linking claims and other types of data—particularly laboratory results data—can substantially expand the range of medication errors that may be detected. For example, it may be possible to assess what proportion of patients exposed to a certain medication had a serum creatinine above a specific level at the time of their initial exposure. Nonetheless, the lack of detailed clinical data can represent a problem.

Chart review can provide valuable additional clinical information, and can be done to supplement claims studies or studies that link claim information with laboratory information. With the chart, it is possible to understand the clinical context, for example the indication for starting a medication, which can sometimes be inferred but rarely determined with certainty with more limited data sources. Chart review is time consuming and expensive, however.

Electronic medical records can provide the clinical detail available with paper-based chart review, but often at much lower cost. It is also possible to search electronic medical records for specific diagnoses, laboratories, and key words suggesting the presence of outcomes. Such records are only used in a minority of outpatient offices today, but they have become the standard in many other countries in primary care, for example the UK. It will be possible to use these records to detect both medication errors and adverse drug events at much lower costs than was previously possible.

THE FUTURE

The future of pharmacoepidemiologic research will include large databases that allow linking of prescription information with clinical and claims data. These types of databases will facilitate the studies of medication errors and adverse drug events. They will also be critical for detecting rare adverse drug events. Sources of data for these databases will include systems of computerized physician order entry, computerized pharmacy, barcoding, and pharmacy benefit managers. Standardized coding of data, that is, the uniform coding of drug names, as well as doses and concentrations, will be an important advancement to allow easy analysis.

Other important issues that must be addressed are representing prescriptions in ways that allow determination of total daily dose, joint documentation of prescriptions and dispensing data to allow determination of patient adherence, clear documentation of conditions like pregnancy or weights of pediatric patients, and improved coding of allergies.

CASE EXAMPLE 27.8: ADVERSE DRUG EVENTS IN OUTPATIENTS

Background

- Most of the data on the frequency of adverse drug events have come from the inpatient setting, and many outpatient studies have relied on chart review or claims data to detect ADEs. Gandhi's 2003 study on the frequency of adverse drug events in a community-living population used a different approach.

Issue

- The goal of the study was to assess the frequency of adverse drug events in an ambulatory primary care population.

Approach

- The frequency of adverse drug events was assessed by calling patients after a visit at which medications were prescribed, to determine whether or not an adverse drug event had occurred, and in addition to review the chart at 3 months.

Results

- Adverse drug events occurred at a rate of 20.9 per 100 patients.
- About eight times as many adverse drug events were identified by calling patients as by reviewing charts.
- While the severity of the ADEs overall was fairly low, about a third were preventable, and 6% were both serious and preventable.

Strength

- The key strength of this approach was that, by calling patients, it was possible to identify many adverse drug events that were not noted in the chart.

Limitation

- The key weakness of this approach is that many of the effects that patients attributed to their medications

(Continued)

may not have been due to the medications at all, but due to other things such as their underlying conditions. The authors attempted to address this by asking the patient's physician in each instance whether they believed that the symptoms related to the medication.

Summary Points

- Calling patients—though expensive and time-consuming—identifies many adverse drug events that are not identified through chart review.
- Almost none of the visits was associated with an ICD-9 code suggesting the presence of an ADE, suggesting that claims data should not be used to estimate the frequency of ADEs of all types in the outpatient setting.
- More work is needed to facilitate assessment of whether a specific patient complaint is related to a medication.

HOSPITAL PHARMACOEPIDEMIOLOGY

Early studies performed in the hospital setting in the 1960s provided much of the initial experience in pharmacoepidemiology. These efforts tracked the drugs administered to patients during their entire hospital stay and systematically recorded adverse events possibly linked to these exposures. The task of actually implementing these activities was not easy in a time before magnetic and electronic storage of data and automated processing became available.

These efforts were followed by the creation of several inpatient databases, some quite comprehensive (e.g., IHS, Medimetrik) and some more selective (e.g., Brigham and Women's Hospital). Creation of these databases was made possible by the availability of computers capable of storing and processing voluminous data collected on large numbers of patients with large numbers of exposures over long periods of time.

However, it was recognized that much of drug exposure was in the outpatient setting, and interest was growing in studying the effects of drugs in that setting. In time, with the increasing availability of claims databases and more recently medical record databases, research has shifted to the larger drug-exposed populations in the outpatient setting. This interest has dominated pharmacoepidemiologic research for several decades. However, a need for data on drug use and drug-related events in the hospital persists.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Volume and Characteristics of Hospital Admissions

A substantial proportion of the medical care provided in the US or elsewhere is in a hospital. In the US in 2002, there were 33.7 million hospital discharges (excluding newborn infants), or a hospital discharge rate of 1174.6 per 10 000 population, with an average length of stay of 4.9 days. In that same year, 31% of total national health spending was for hospital care, totaling \$486.5 billion, and 12% of all US drug sales were in hospitals.

Variations in hospital care outcomes can be ascribed, to some degree, to the characteristics of the hospitals themselves as much as the characteristics of the patients admitted to them. For example, one study found that older patients who needed high-risk cardiovascular or cancer surgeries were more likely to survive in hospitals that performed a high volume of these complicated surgical procedures as compared to patients admitted to hospitals with little experience with these surgeries. Another study found fewer deaths among elderly Medicare patients hospitalized for first-time heart attack when staffing levels of registered nurses (RNs) were higher as opposed to substituting licensed practical nurses (LPNs) for RNs.

Characteristics of Hospitalized Patients and Hospital Drug Use

Hospitalized patients receive multiple drugs during their hospital stay, which is mostly warranted because these patients are older, sicker, and have multiple concurrent diseases.

Almost 40% (12.7 million) of all discharges from short-stay hospitals in 2002 involved patients aged 65 years and older. Further, the elderly are more likely to suffer complications from their care and to die.

The practice of polypharmacy introduces the risks of unintended drug-drug interactions, prescribing suboptimal drug regimens, and inadequate laboratory monitoring of drug therapy. It is well documented that patients receiving multiple drugs during their hospital stay have higher rates of drug reactions.

Another aspect of relevance to pharmacoepidemiologic research is the observation that most drug reactions tend to occur in the first five days on a drug; therefore, active surveillance of drug use during this period is likely to detect most adverse reactions.

These considerations illustrate that patient and hospital characteristics and practices need to be incorporated in study

designs, analyses, and interpretations of data collected from hospital settings.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGIC RESEARCH

Logistical Issues

Developing *complete information* on total drug exposure during a hospital stay is a major challenge for pharmacoepidemiologic research in the hospital setting, where patients are administered drugs at multiple sites, such as emergency rooms, operating rooms, radiology, and patient room beds. Drugs are also administered by multiple types of personnel, e.g., nurses, physicians, radiology technicians, etc. Further, these administrations can be recorded in multiple different forms, such as progress notes, pharmacy records, and nursing medication administration records. The task of measuring hospital drug exposure fully and accurately can therefore be daunting. Yet, without complete information about inpatient drug exposure, inferences about adverse drug reactions associated with particular exposures are impossible to make credibly.

Presumably, as hospitals increasingly implement comprehensive, linked, automated inpatient data systems, the capture of complete inpatient drug exposure will be addressed. Yet, even in hospitals that made the investment to adopt, for example, a Computerized Prescriber Order Entry (CPOE) system, many gaps remain in the completeness of information that could be useful and necessary for research purposes. CPOE data can be incomplete due to omitted high-risk populations, such as neonates or patients receiving chemotherapy; inability to handle selected complex clinical situations, e.g., patient-controlled analgesia and operating room medications; and CPOE systems that are not fully integrated with other medication system components, e.g., barcode point-of-care systems. Also, many CPOE systems are stand-alone products, without an interface with other institution-wide systems.

Methodological Issues

Besides the problem of measuring total exposure, there is the issue of uncertain *validity of the drug information* in the hospital medical record. Drugs dispensed from a pharmacy are not always administered to patients, actual drug administrations are not always recorded, and drugs ordered for one patient might be dispensed to a different patient. Such errors of omission or commission need to be explored

and assessed before relying on the hospital medical record as the sole source of information to answer questions about drug-associated risks.

The *validity of diagnosis information* in the hospital medical record may also be uncertain. Studies have found that diagnosed conditions, even those that were treated, were not always listed, and that procedures not performed in the operating room were often omitted from the chart. Of course, with electronic data entry medical records and greater linking of the pharmacy and clinical databases, this problem of incomplete or erroneous hospital data may diminish, though it is unlikely to be completely eliminated.

Another concern is the absence of *inpatient information linked to outpatient information*. An adverse event occurring in a hospital within a few days after admission may be a reaction to a drug used before admission. Similarly, an adverse event occurring after discharge may be due to a drug exposure during the hospital stay. Yet, these associations may be missed with inpatient data unless there is linkage to prior and follow-up data.

Another methodological problem is the *uniqueness of drug exposure* in the hospital setting. Hospitals generally use a restricted set of drugs as specified by their formularies. This limits the types of drugs that can be studied for adverse reactions. Another limitation for research is that hospitals use some drugs, dosages (often higher), and forms of administration (e.g., parenteral) that differ from those used in outpatient care. Also, drug usage in the hospital tends to be short-term in contrast to long-term drug use by outpatients. Further, the drugs used differ from hospital to hospital. The uniqueness of these patterns of inpatient drug use may have implications for the generalizability of findings from studies in particular hospitals. Also, because of small populations or the infrequent use of some drugs, it may not be possible to detect the rare adverse reactions to most drugs.

Another methodological problem is that because hospitalized patients are sicker than other people taking prescribed drugs, they are more likely to receive multiple drugs, and, consequently, more likely to experience an adverse drug event. However, sicker patients tend to have more underlying medical problems, a fact that makes it more difficult for a physician to discern an adverse reaction to a drug from an event due to another cause. Patients may react differently to the same drug, depending on various comorbidities, demographic profiles, and personal histories, thereby further confounding a possible association between exposure and event.

Furthermore, because hospitalized patients tend to experience many events during the course of their stay, there may be a *tendency to record only the most extreme or dramatic*

events. Therefore, certain mild adverse drug events (ADEs) may not be completely recorded, and would not be available for studies of drug effects.

When using hospital patients for a pharmacoepidemiology study, one must be mindful also of the potential for *referral bias*. If an exposure is related to likelihood of hospitalization, then outcomes of hospitalization may falsely appear related to that exposure. Relatedly, variations in hospital practices, such as ordering diagnostic or laboratory tests for patients on drugs, the effects of which are monitored, may produce an *ascertainment bias* and, with that, abnormally high or low rates of apparent ADEs.

It is important to note that the definition of an *appropriate denominator for calculations of adverse drug reactions* depends on the question of interest. The total quantity of drug received is the more useful denominator if one is interested in the incidence rate of adverse reaction for a specific drug. The total patient population is the appropriate denominator if one is interested in the overall rate of adverse drug reactions in a hospital. The total number of patients exposed to a drug may be the most appropriate denominator, but is not always available.

A different methodologic problem emerges from problems in *hospital staff participation*. Hospitals are complex organizations with large numbers of organizational units and large numbers of personnel who function under stressful conditions in response to complex medical problems. Approaching these busy professionals with requests to identify and refer potentially eligible patients to the research team is an imposition that they tend to resist. It requires much effort on the part of the researcher to influence hospital physicians to agree to participate in an epidemiologic study and permit contacting their patients. It also requires lengthy negotiation and creative arrangements to get access to hospital admission logs, surgical logs, or patients' medical records. Concerns about privacy, of course, complicate the whole endeavor.

Finally, hospitals are huge bureaucracies obviously not created for research purposes. They collect vast amounts of information that can serve as rich resources for research but the information is collected for clinical care and administrative purposes rather than research purposes. In other words, *medical records are not organized for research*, i.e., in a fashion that makes it simple for an investigator to review and abstract the medical record for a particular research study. The medical record may not even contain certain information useful for an epidemiologic study (e.g., diet, exercise, occupational history). Missing information on important potential confounders may limit inferences about drug effects.

On the positive side, the study of inpatient drug effects is enhanced by the fact that there is a large amount of information recorded due to continuous monitoring and measurement of patients' conditions. Furthermore, there is less likely to be a problem of patient compliance with drug intake; therefore, the connection with a subsequent adverse reaction is more credible.

EXAMPLES OF CURRENTLY AVAILABLE SOLUTIONS

Intensive Hospital-Based Surveillance

In Scotland, Finney proposed in 1965 a framework for monitoring drug use and drug reactions that consisted of routine recording of demographic and clinical information on hospitalized patients, including all drugs administered throughout their hospital stay. Then, by comparing the rates of events occurring in these patients and performing cohort studies, one could detect adverse reactions, whether or not physicians suspected any associations between drugs and events.

Concurrently, an intensive inpatient drug monitoring program was developed at the Johns Hopkins University. The cohort under surveillance was large—as it had to include a large at-risk population to be able to detect infrequent adverse reactions—as was the amount of information systematically collected and recorded on patients in the cohort. There have been several other notable series of intensive hospital-based cohort studies, including the Shands Hospital in Florida, the Comprehensive Hospital Drug Monitoring Berne, in Switzerland, and, particularly, the Boston Collaborative Drug Surveillance Program.

Ultimately, the most comprehensive intensive in-hospital drug surveillance program—started in 1966 and accumulating information on over 50 000 medical inpatients over a period of almost 20 years—was the Boston Collaborative Drug Surveillance Program (BCDSP). After initially monitoring the medical patients of a single Boston hospital, it was expanded to include multiple hospitals in several states in the US and abroad. Collection of new data from medical wards was discontinued in 1977, and collection of data from surgical wards (on about 5200 patients) was discontinued in 1984, but this old database remains available and has generated numerous research papers about older drugs. A similar intensive drug surveillance program was later launched by these investigators in pediatric wards at several teaching and community hospitals in several US states, accumulating data on 10 297 pediatric patients.

Its early approach was similar to that at Johns Hopkins, that is, continuously tracking and monitoring drug usage

and adverse events of hospital patients. Later they added an inquiry about drug exposure in the 3 months before admission. Collaborative arrangements with multiple hospitals involved hiring local nurses at each site and training them to follow standardized procedures for collecting information from consecutive admissions to the hospital. Detailed information was collected shortly after admission by structured patient interviews. Then, detailed information about drugs administered during hospitalization was obtained through chart reviews. Also, the nurse monitors attended ward rounds and communicated with attending physicians to obtain their judgments about suspected adverse drug-related events. Further, in order not to miss associations that might become the subject of later testing, a few selected major events (such as sudden death, jaundice, renal failure, gastrointestinal bleeding, and psychosis) were targeted for recording, regardless of whether they were thought to be drug related. All data collected were submitted to the Boston central research office to be subjected to tests for accuracy and completeness and to be processed and analyzed.

Because patient interviews focused on drugs used before hospitalization in relation to the causes of admission to the hospital, this database made it possible to examine and control for drug usage preceding hospitalization. Also, because medical record data in the BCDSP were augmented with data from patient interviews that asked about alcohol and tobacco use, this database made it possible to control for these important confounders of drug effects. Another advantage of the BCDSP was its large size, permitting the study of rare adverse medical events. Also, with multiple countries contributing data, study results could be evaluated for consistency cross-nationally. However, the major component of BCDSP data (for medical inpatients) is now almost 30 years old, and cannot support studies of the many important new drugs that have become available since the mid-1970s. Nonetheless, the BCDSP has supported a wide range of important studies since its inception.

Inpatient Databases

Multisite Databases

The Medimetrik and IHS databases were examples of early multisite hospital databases started as commercial enterprises. The Medimetrik database, for example, discontinued by 1988, could support research on rare drug events because of its large size, as data were contributed from administrative, pharmacy, and discharge sites at 50 hospitals. However, it was not commercially viable, given the large costs of gathering such data.

One noncommercial, large data collection system in a hospital is described below.

Brigham and Women's Hospital

Brigham and Women's Hospital (BWH), a 720-bed medical center in Boston, has developed a computerized data collection and reporting capability called the Brigham Integrated Computing System (BICS) into which all orders for medications, laboratory tests, and other therapeutic interventions for all adult inpatients are entered. This database contains complete patient demographic information, discharge diagnoses, procedures, laboratory results, and pharmacy data from 1987, with less complete information available beginning in 1981.

The BICS provides BWH with almost all of its clinical, administrative, and financial computing services. The BICS clinical information system includes a wide array of services such as test results review, longitudinal medical records, provider order entry, critical pathway management, critical-event detection and altering, automated inpatient summaries, operating-room scheduling, coverage lists, and an online reference library.

Examples of the ways in which BICS is used to contribute to improved inpatient care and cost savings include a computerized physician order entry designed to display drug use guidelines, offer relevant alternatives, and suggest appropriate doses and frequencies whenever physicians enter drug orders for their patients; a computer intervention designed to notify physicians when inpatients remain on expensive intravenous medications after they become able to take bioequivalent oral alternatives; another computer intervention designed to provide physicians with electronic notification that certain clinical laboratory tests may be redundant; a computer application to protect against errors in chemotherapy ordering and dosing and to coordinate the outpatient and inpatient chemotherapy services; and a computer-based ADE monitor.

A study by Teich *et al.* (2000) showed that computerized guidelines prompted physicians to increase the use of recommended drugs and decrease the proportion of doses exceeding the recommended maximum. A study by Bates *et al.* (1999) found that computerized alerts about potentially redundant clinical laboratory tests were effective, but limited because many tests were performed without the corresponding computer reminders, and many orders were not screened for redundancy. A study by Jha *et al.* (1998) showed that using a computer-based ADE monitor resulted in identification of fewer ADEs than did chart review, leading them to conclude that different detection methods capture different events because of minimal overlap among the ADEs identified by the different methods.

Advantages and Limitations

There is merit and utility in the efforts of hospitals to establish comprehensive integrated automated data systems for enhancing the quality of patient care, reducing medical errors and adverse reactions to treatment, and supporting research. (See Case Example 27.9.) From the perspective of pharmacoepidemiologic research, however, it is unfortunate that these inpatient data systems are freestanding instead of integrated across institutions. The size and composition of patient populations in individual hospitals are likely to limit the ability to detect events of low incidence and produce generalizable results.

New Hospital-Based Adverse Drug Reaction Monitoring and Drug Use Evaluation Programs

A series of initiatives by the US Joint Commission on Accreditation of Healthcare Organizations (JCAHO)—the private nonprofit organization that must provide a hospital accreditation for it to qualify for payment by Medicare without separate inspections from the Federal government—have contributed significantly to the further development of hospital pharmacoepidemiology. These JCAHO initiatives began in 1986 with the “Agenda for Change,” which emphasized the quality of hospital performance. This outlined the components of hospital drug use and associated requirements, specifically the development of adverse drug monitoring and reporting programs and drug usage evaluation programs. This was followed in 1989 by the “National Forum on Clinical Indicator Development,” which encouraged the creation of expert panels and data-driven monitoring systems and outcome assessment. In 1990, JCAHO established the “Medication Use Task Force,” which developed specific, quantifiable indicators of good drug use with the intent that data on these indicators would be provided by all hospitals, and then a hospital would be provided with data on where it stood versus other hospitals. These indicators were never implemented. The broad goals of these and related programs were: to improve the quality of patient care by improving the clinical use of medications and minimizing adverse drug reactions; decrease hospital costs by eliminating inappropriate use of drugs or offering acceptable low cost substitutions; and decrease liability associated with the inappropriate use of high-risk drugs.

As an example of a response by hospitals to the JCAHO initiatives, the Hospital of the University of Pennsylvania (HUP) established the Drug Use and Effects Committee (DUEC), as a subcommittee of the Pharmacy and Therapeutics Committee, to provide guidance and oversight over these programs. To fulfill its mission, the *Adverse Drug Event*

(ADEs) *Surveillance Program* at HUP developed operational definitions of ADEs, publicized the program to nurses and physicians, arranged a mechanism for communicating spontaneous reports of ADEs, screened computerized discharge diagnoses for ADEs, targeted drugs, patients, and sites for intensive follow-up, triaged reports for in-depth investigation and reporting to the FDA, tabulated and analyzed the accumulated data on ADEs, reviewed these results on a regular basis, and set up plans for any necessary follow-up. All of these activities have been routinized and are continuously ongoing. Further, regular reports identify the top medications involved in ADEs, the severity of the reaction (mild or serious), whether they occur in the inpatient or outpatient setting, whether the event was dose-related, idiosyncratic, attributed to a new drug (defined as available on the market for three years) or multiple medications, and the source that reported the ADEs (e.g., admissions list, pharmacists, physicians, pharmacy students, tracer drugs, lab signals). Most of the tracer drugs are antidotes, and lab signals include all critical lab results reported in the hospital. The primary focus is on documenting severe, preventable, and rare ADEs, with secondary emphasis placed on documenting non-serious ADEs that are not dose-related.

As part of monitoring the program on adverse drug reactions, the DUEC also reviews monthly reports of *pharmacist interventions*. These can involve *preventive measures* (such as blocking an order for a prescription if the dose or regimen is considered inappropriate or because of potential drug–drug interactions) as well as *cost avoidance measures* (such as identifying unnecessary drug use when a physician orders two drugs when one drug may be enough to achieve a therapeutic goal, suggesting that the regimen could be decreased, or suggesting a lower cost substitute drug). Along with these reviews, the DUEC’s *Drug Usage Evaluation (DUE) Program* developed criteria for appropriate drug use, used the pharmacy computer to identify exposed patients, developed data entry forms for chart abstracting, performed medical chart reviews, organized and analyzed the collected data, reported the results to a review committee, and designed and implemented interventions to address any problems identified. Again, all these activities are now routinized and ongoing.

DUEC’s *Cost Containment Program* was developed to identify potential problem areas based on ADEs, DUEs, and/or drugs generating the most expense, and to intercede to reduce costs, either through educational campaigns, formulary modifications, or other interventions.

In the years since its establishment at HUP, the contributions of the DUEC to patient safety can be discerned from reviewing a few examples of its activities and outcomes.

Table 27.5. Medications involved in ADEs: Drug Use and Effects Committee, Hospital of the University of Pennsylvania, 2004

Medication	Total (%)
Warfarin	26.4
Heparin	17.4
Insulin	14.9
Acetaminophen	8.0
Amiodarone	7.0
Cefepime	5.5
Multiple medications	5.5
Phenytoin	5.5
Vancomycin	4.5
Amphotericin B	3.0
Alemtuzumab	2.5

Table 27.5 and Figures 27.2 and 27.3 present some data from the DUEC ADE program. Included in the reports have been several previously unrecognized ADEs, which have led to changes in drug labels or manufacturing. As another example, the DUEC requested a DUE on allergy to determine compliance with the hospital's policies concerning the documentation of patient height, weight, and allergy information and the consistency of this information between sources. The findings from this DUE analysis triggered several recommendations by the DUEC, including: (i) developing a policy concerning which health care professional should be responsible for collecting and disseminating height, weight, and allergy information; (ii) incorporating height, weight, and allergy information into pharmacist medication histories; (iii) instituting data entry prompts for missing height, weight, and allergy information into both the computerized physician order entry system and the pharmacy computer system; (iv) educating health care professionals on the importance of recording

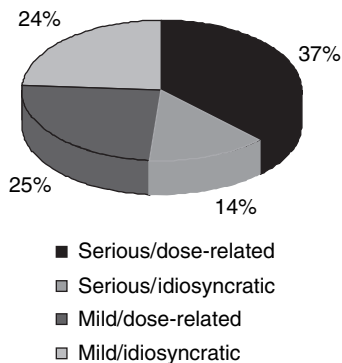


Figure 27.2. Adverse drug events, 2003—Hospital of the University of Pennsylvania.

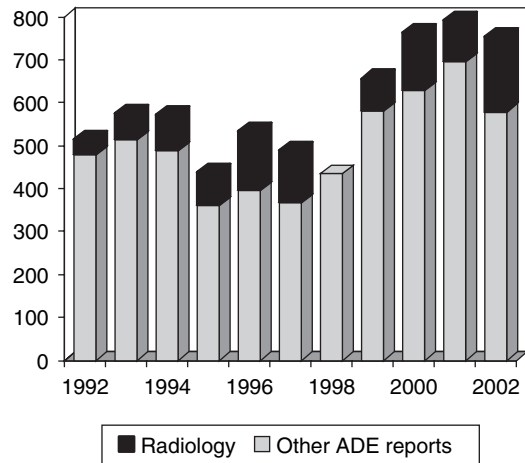
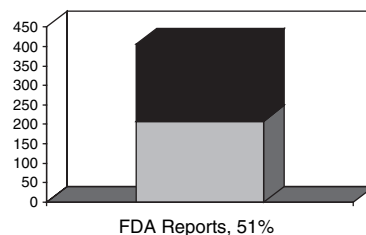


Figure 27.3. Adverse drug events, 1992–2002—Hospital of the University of Pennsylvania.

reaction information on allergies; and (v) making patient allergy information a mandatory field before adding medication orders. Specific actions that were undertaken involved: (i) organizing a meeting with Admissions to discuss the policy and expectations of capturing the patient height, weight, and allergy history; (ii) investigating solutions with Information Systems regarding the capability of the Clinical Manager computer system to capture height, weight, and allergy information; and (iii) following up with Information Systems regarding pharmacist access to the computerized physician order entry system to add or change allergy information.

The DUEC's oversight of the inpatient Cost Containment Program has targeted, for example, the antibiotic management program and anticoagulation management



program. Some of these new initiatives involve including the daily cost information for antibiotic therapy on the Microbiology Laboratories Sensitivity report and including in the formulary the cost data for drug classes in table format to allow for easy reading by prescribers.

Obviously, having a committee of experts regularly reviewing detailed information on drug use and ADEs in the hospital provides for continuous quality assurance and enhanced patient safety. However, in addition to these monitoring activities, the DUEC also designs and sponsors interventions.

THE FUTURE

According to all current indications, the trend toward more automated, comprehensive inpatient data collection and storage is expected to continue in the future. This trend has been accelerated by requirements by the JCAHO, the drive for improving patient safety, and the data indicating that the application of information technology may assist in achieving this goal. Nonetheless, further improvements in inpatient data are needed.

First, inpatient data systems need to be dynamic so as to change as new medical developments occur. For example, as new drugs are approved for use and genetic markers for diagnostic purposes are introduced, the data systems must be expanded on an ongoing basis to include such developments. Likewise, data systems must capture new ADEs as they emerge

Second, inpatient data systems need to further develop and improve in their capability to extract free text from clinical notes throughout the medical record. This is necessary to create a fully automated medical record for patient care and research purposes that can be linked to other elements in the data system such as drug use.

Third, inpatient data systems need to be developed and implemented in long-term care and nursing home facilities where a large proportion of drugs are used in elderly patients that are not currently well surveyed. The creation of such data systems will not only provide for improved patient care in these facilities, but will also make possible the linking of hospital and long-term care databases to permit tracing and follow-up of patients transferring between these types of inpatient care facilities.

Fourth, there still remains a need for the development of regional automated data systems that link inpatient data systems from multiple hospitals and link with outpatient data as well. For patient treatment, this will provide for more coordinated care as patients move between different providers

and levels of care. For research purposes, regional data systems will provide a defined population, increased sample size, and longitudinal information for individual patients.

Fifth, a further need is to extend linkages of inpatient data systems across a nationally representative sample of hospitals to create a truly powerful resource for medical research.

Sixth, despite the prospects for much improved inpatient data systems, privacy regulations severely restrict access to these data for research purposes.

Thus, in conclusion, the future is likely to see a continuation of the increased interest in hospital pharmacoepidemiology that has emerged in the past few years, greatly accelerated by the increasing computerization of hospital care. In the process, we will hopefully be able to learn more, and faster, about those drugs we use primarily in hospitals.

CASE EXAMPLE 27.9:

NON-EXPERIMENTAL STUDY OF INPATIENT DRUG EFFECTS: PARENTERAL KETOROLAC

Background

- Case reports suggested that parenteral use of ketorolac might be associated with multiple different serious adverse effects. As a parenteral drug, however, its use is primarily in hospitals.

Question

- What are the side effects associated with the use of parenteral ketorolac?

Approach

- There is no database that contains sufficient numbers of hospitalized patients to answer this question.
- A retrospective cohort study was performed, comparing 9907 inpatients given 10 279 courses of parenteral ketorolac to 10 248 inpatients given parenteral narcotics and no parenteral ketorolac, matched on hospital, admission service, and date of initiation of therapy.
- Data were collected by computer-assisted chart review, within 35 hospitals.

Results

- Parenteral ketorolac was indeed associated with an increased risk of gastrointestinal bleeding, especially in the elderly, with a dose–response relationship, but

only in those who used the drug for a prolonged time period.

- Parenteral narcotics had a higher risk of respiratory depression.
- Overall, the risk/benefit balance of parenteral ketorolac versus parenteral opiates was deemed to be similar, but limiting the use of ketorolac (e.g., duration <5 days) would improve risk/benefit balance further.
- The choice of the optimal drug needs to be made on a patient-specific basis.

Strengths

- Concurrent and comparable control group.
- Huge sample size.
- Natural setting.

Limitations

- Risk of selection bias: the different analgesics were used in patients who were somewhat different.
- Expensive, time-consuming study.

Summary Points

- Outcomes that are relatively uncommon require large sample sizes.
- The conduct of large studies of drugs used in hospitals requires at this point *ad hoc* data collection.
- A system to link computerized data collected by multiple hospitals would be a useful tool in the future.

CONCLUSIONS

In this chapter we have presented many special applications of pharmacoepidemiology. Each application uses pharmacoepidemiology approaches, and each has unique methodologic problems, in answering its type of special clinical question. In each case, there are solutions being used and future developments that are still needed. Hopefully, the comparisons among them are useful in and of themselves, as well as being useful to display the breadth of applications and approaches used in this field.

Key Points

Studies of Drug Utilization

- Drug utilization studies can be performed to quantify and identify problems in drug utilization, monitor changes in utilization patterns, or evaluate the impact of interventions.
- Drug utilization studies may be conducted on an ongoing basis in programs for improving the quality of drug use.
- Assessing the appropriateness of drug utilization requires data on indication for treatment, patient characteristics, drug regimen, concomitant diseases, and concurrent use of other medications.
- When assessing quality of care, drug utilization studies must often rely on multiple sources of data.

Evaluating and Improving Physician Prescribing

- Quality problems in prescribing exist at the level of medication overuse (e.g., antibiotics for viral respiratory tract infections in adults and children), misuse (e.g., coxibs for patients at low risk of upper gastrointestinal bleeding or high risk of cardiac events), and underuse (e.g., bisphosphonates in the secondary prevention of osteoporosis-related fractures or inhaled corticosteroids for reactive airways disease).
- The vast majority of prescribing problems are related to underuse of proven effective treatments.
- Passive interventions, such as dissemination of printed or emailed guidelines, drug utilization reviews and medication profiles, or traditional didactic continuing medical education lectures, are unlikely to improve practice.
- More active intervention strategies (e.g., point-of-care reminders, educational outreach, achievable benchmarks with audit and feedback), especially when combined together to overcome barriers at the level of the system, the physician, and the patient, are able to modestly improve the quality of prescribing.

(Continued)

- Just as with the adoption of drugs and devices, interventions to improve physician prescribing need to be tested in rigorous controlled trials before widespread and expensive implementation. In particular, investigators should consider “mixed-method” evaluations (i.e., quantitative data, prescriber surveys, qualitative inquiry about barriers and facilitators to adoption) of their interventions to better understand what works or does not work—and why.

Drug Utilization Review

- Although there is evidence that some forms of drug utilization review can modestly impact prescribing, evidence that it improves clinical outcomes is lacking.
- Drug utilization review programs can have unintended consequences.
- Drug utilization review programs should be subjected to rigorous evaluation.

Special Methodologic Issues in Pharmacoepidemiologic Studies of Vaccine Study

- There are still substantial gaps and limitations in our knowledge of many vaccine safety issues.
- A high standard of safety is required for vaccines due to the large number of persons who are exposed, some of whom are compelled to do so by law or public health regulations.
- New research capacity, such as the Vaccine Safety Datalink, provides powerful tools to address many safety concerns.

Pharmacoepidemiologic Studies of Devices

- Medical devices are of public health importance.
- Medical devices and their use have diverse characteristics and are complex.
- Existing data sources have limited utility for medical device epidemiology because complete documentation of device use is not routine.
- Another barrier to good documentation of medical device use is the lack of a detailed identification system (analogous to the National Drug Code) for medical devices.
- Although some excellent epidemiology studies have been reported, barriers to the conduct of efficient and robust studies need to be removed.

Studies of Drug-Induced Birth Defects

- Unlike most drug effects, human teratogenesis can only be identified in the postmarketing setting.
- There is remarkable ignorance about the teratogenic risk/safety of the vast majority of prescription and non-prescription drugs taken by pregnant women.
- Most human teratogens increase risks of specific defects (not birth defects overall).
- Pregnancy registries are well suited to identify “high-risk teratogens” (e.g., thalidomide, isotretinoin), but do not have power to identify “moderate risk teratogens”; case–control approaches have the statistical power necessary to identify “moderate-risk teratogens” and to identify relative safety.
- Combining the complementary strengths of cohort and case–control approaches can provide a comprehensive design to identify human teratogens and to establish ranges of risk/safety for the wide range of medications taken by pregnant women.

Pharmacoepidemiology and Risk Management

- Risk management encompasses those strategies intended to shift the balance of benefit and risk for a drug to an acceptable level.
- To date, most strategies have been found to be of no or marginal effectiveness.
- Additional methods are needed that effectively optimize the balance of benefit and risk and that succeed in modifying physician prescribing behavior.

The Use of Pharmacoepidemiology to Study Medication Errors

- Medication errors are very common, compared to adverse drug events, and relatively few result in injury.
- In most studies, about a third of adverse drug events have been preventable.
- The epidemiology of medication errors and adverse drug events has been fairly well described for hospitalized adults, but less information is available for specific populations, and for the ambulatory setting.
- It is possible now to detect many medication errors using large claims databases, and as it becomes possible to link these data with more types of clinical data, including especially laboratory and diagnosis data, it will be feasible to more accurately assess the frequency of medication errors across populations.
- The increasing use of electronic health records should have a dramatic effect on our ability to do research in this area using pharmacoepidemiologic techniques.

Hospital Pharmacoepidemiology

- A substantial proportion of the medical care provided is in a hospital, treating very ill patients with multiple simultaneous drugs, some more toxic than drugs used in outpatients.
- While pharmacoepidemiology began as a field with hospital-based data, recently developed automated databases rarely include drugs administered in hospitals.
- Hospitals are remarkably complex organizations, and drug use throughout a hospital is equally complex, administered by multiple types of personnel and recorded in multiple different medical records forms. The task of measuring hospital drug exposure fully and accurately can therefore be daunting.
- Intensive hospital-based surveillance consisted of routine prospective recording of demographic and clinical information on hospitalized patients, including all drugs administered throughout their hospital stay. Then, by comparing the rates of events occurring in these patients and performing cohort studies, one could detect adverse reactions, whether or not physicians suspected any associations between drugs and events.
- Hospitals now have *ad hoc* adverse drug reaction monitoring and drug use evaluation programs.

DISCLAIMER

The views expressed are those of the authors and do not necessarily represent those of the US Food and Drug Administration.

SUGGESTED FURTHER READINGS

STUDIES OF DRUG UTILIZATION

- Brown TR. *Handbook of Institutional Pharmacy Practice*, 4th edn. Bethesda, MD: American Society of Health System Pharmacists, 2006.
- Dartnell JGA. *Understanding, Influencing and Evaluating Drug Use*. Australia: Therapeutic Guidelines Limited, 2001.
- Dukes MNG. *Drug Utilization Studies: Methods and Uses*. *European Series No. 45*. Copenhagen: World Health Organization, Regional Office for Europe, 1993.
- 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.

- Kauffman 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.
- Kidder D, Bae J. Evaluation results from prospective drug utilization review: Medicaid demonstrations. *Health Care Financ Rev* 1999; **20**: 107–18.
- Mino-León D, Figueras A, Amato D, Laporte J-R. Treatment of type 2 diabetes in primary health care: a drug utilization study. *Ann Pharmacother* 2005; **39**: 441–5.
- Rucker TD. Data, sources, and limitations. *JAMA* 1974; **230**: 888–90.
- Schiff GD, Rucker TD. Computerized prescribing: building the electronic infrastructure for better medication usage. *JAMA* 1998; **279**: 1024–9.
- World Health Organization. *How To Investigate Drug Use in Health Facilities: Selected Drug Use Indicators*, WHO/DAP/93.1. Geneva: World Health Organization, 1993.
- 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.

EVALUATING AND IMPROVING PHYSICIAN PRESCRIBING

- 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.
- 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.
- 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.
- 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.
- Donner A, Birkett N, Buck C. Randomization by cluster—samples size requirements and analysis. *Am J Epidemiology* 1981; **114**: 906–14.
- 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.
- Eisenberg JM. *Doctors' Decisions and the Cost of Medical Care*. Ann Arbor, MI: Health Administration Press Perspectives, 1986.
- 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.
- Greco PJ, Eisenberg JM. Changing physicians' practices. *N Engl J Med* 1993; **329**: 1271–4.
- Green LW, Kreuter MW. *Health Promotion Planning: An Educational and Environmental Approach*. Mountain View, CA: HJ Kaiser Foundation, 1991.
- 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.
- 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.
- 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.
- 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.
- Lee TH. A broader concept of medical errors. *N Engl J Med* 2002; **347**: 1965–7.
- 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.
- 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.
- Rousseau N, McColl E, Newton J, Grimshaw J, Eccles M. Practice based, longitudinal, qualitative interview study of computerized evidence based guidelines in primary care. *BMJ* 2003; **326**: 314–22.
- 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.
- Soumerai SB, Avorn JL. Principles of educational outreach ("academic detailing") to improve clinical decision making. *JAMA* 1990; **263**: 549–56.
- 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.
- Tamblyn R, Huang A, Perreault R, Jacques A, Roy D, Hanley J *et al*. The medical office of the 21st century (moxxi): effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. *Can Med Assoc J* 2003; **169**: 549–56.
- Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N *et al*. Educational outreach visits: effects on professional practice and health care outcomes (Cochrane Review). In: *The Cochrane Library*, Issue 4. Chichester: John Wiley & Sons, 2003.

DRUG UTILIZATION REVIEW

- Hennessy S, Bilker WB, Weber A, Zhou L, Brensinger C, Wang Y, Strom BL. Retrospective Drug Utilization Review, Prescribing Errors, and Clinical Outcomes. *JAMA* 2003; **290**: 1494–99.
- Smith DH, Christensen DB, Stergachis A, Holmes G. A randomized controlled trial of a drug use review intervention for sedative hypnotic medications. *Med Care* 1998; **36**: 1013–21.
- Soumerai SB, Lipton HL. Computer-based drug-utilization review—risk, benefit, or boondoggle? *N Engl J Med* 1995; **332**: 1641–5.

SPECIAL METHODOLOGIC ISSUES IN PHARMACOEPIDEMIOLOGY STUDIES OF VACCINE SAFETY

- 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.
- Baylor NW, Midthun K. Regulation and testing of vaccines. In: Plotkin S, Orenstein WA, eds. *Vaccines*, 4th edn. Philadelphia, PA: W.B. Saunders, 2003; pp. 1539–81.
- 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.

- 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.
- Davis RL, Kolczak M, Lewis E, Nordin J, Goodman M, Shay DK, Platt R, Black S, Shinefield H, Chen RT. Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology* 2005; **16**: 336–41.
- DeStefano F. The Vaccine Safety Datalink project. *Pharmacoepidemiol Drug Saf* 2001; **10**: 403–6.
- 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.
- Fine PE, Chen RT. Confounding in studies of adverse reactions to vaccines. *Am J Epidemiol* 1992; **136**: 121–35.
- 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.
- Halsell JS, Riddle JR, Atwood JE, Gardner P, Shope R, Poland GA *et al.* Myopericarditis following smallpox vaccination among vaccinia-naïve US military personnel. *JAMA* 2003; **289**: 3283–9.
- Horton R. A statement by the editors of The Lancet. *Lancet* 2004; **363**: 820–1.
- Horton R. The lessons of MMR. *Lancet* 2004; **363**: 747–9.
- Howson CP, Howe CJ, Fineberg HV. *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.
- 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**: 410–16.
- Murch SH, Anthony A, Casson DH, Malik M, Berelowitz M, Dhillon AP *et al.* Retraction of an interpretation. *Lancet* 2004; **363**: 750.
- Plotkin SA, Rupprecht CE, Koprowski H. Rabies vaccine. In: Plotkin SA, Mortimer EA, eds, *Vaccines*, 4th edn. Philadelphia: W.B. Saunders, 2003; pp. 1011–38.
- Ray WA, Griffin MR. Re: “Confounding in studies of adverse reactions to vaccines.” *Am J Epidemiol* 1994; **139**: 229–30.
- Schonberger LB, Bregman DJ, Sullivan-Bolyai JZ, Keenlyside RA, Ziegler DW, Retallia 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.
- Weldon D. Before The Institute of Medicine, February 9, 2004. Available at: <http://www.iom.edu/includes/DBFile.asp?id=19029>. Accessed: April 1, 2004.
- Wise RP, Kiminyo KP, Salive ME. Hair loss after routine immunizations. *JAMA* 1997; **278**: 1176–8.
- PHARMACOEPIDEMIOLOGIC STUDIES OF DEVICES
- Bright RA, Jeng LL, Moore RM. National survey of self-reported breast implants: 1988 estimates. *J Long Term Eff Med Implants* 1993; **3**: 81–9.
- Bright RA. Pharmacoepidemiology studies of devices. In: Strom BL, ed., *Pharmacoepidemiology*, 4th edn. Chichester: John Wiley and Sons, 2005.
- Daley WR, Kaczmarek RG. The epidemiology of cardiac pace-makers in the older US population. *J Am Geriatr Soc* 1998; **46**: 1016–19.
- Executive Summary for Aspden P, Corrigan JM, Wolcott J, Erickson SM, eds. *Patient Safety: Achieving a New Standard for Care*. Available from http://books.nap.edu/execsumm_pdf/10863.pdf. Accessed January 2004.
- Feigl DW, Gardner SN, McClellan M. Ensuring safe and effective medical devices. *N Engl J Med* 2003; **348**: 191–2.
- Food and Drug Administration homepage. <http://www.fda.gov>. Accessed February 2006.
- Global Harmonization Task Force homepage. <http://www.ghrf.org>. Accessed February 2006.
- Global Medical Device Nomenclature (GMDN) Homepage. Available from <http://www.gmdn.org/>. Accessed December 2003.
- IMS America. Hospital Supply Index 1998. Available from <http://www.imshealth.com/>. Accessed January 2004.
- Jeng LL, Moore RM, Kaczmarek RG, Placek PJ, Bright RA. How frequently are home pregnancy tests used? Results from the 1988 National Maternal and Infant Health Survey. *Birth* 1991; **18**: 11–13.
- Kurland LT, Molgaard CA. The patient record in epidemiology. *Sci Am* 1981; **245**: 54–63.
- National Center for Health Statistics homepage. <http://www.cdc.gov/nchs>. Accessed February 2006.
- Nationwide Inpatient Sample (NIS): Powerful Database for Analyzing Hospital Care. United States Agency for Healthcare Research and Quality, July 2003. Available from <http://www.ahrq.gov/data/hcup/hcupnis.html>. Accessed January 2004.
- Poggio EC, Glynn RJ, Schein OD, Seddon JM, Shannon MJ, Scardino VA *et al.* The incidence of ulcerative keratitis among users of daily-wear and extended-wear soft contact lenses. *N Engl J Med* 1989; **321**: 779–83.
- 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.
- Schein OD, Glynn RJ, Poggio EC, Seddon JM, Kenyon KR. The relative risk of ulcerative keratitis among users of daily-wear and extended-wear soft contact lenses. A case-control study. *N Engl J Med* 1989; **321**: 773–8.
- Silverman BG, Gross TP, Kaczmarek RG, Hamilton P, Hamburger S. The epidemiology of pacemaker implantation in the United States. *Public Health Rep* 1995; **110**: 42–6.
- Silverman BG, Brown SL, Bright RA, Kaczmarek RK, Arrowsmith-Lowe JB, Kessler DA. Reported complications of

silicone gel breast implants: an epidemiologic review. *Ann Intern Med* 1996; **124**: 744–56.

Spera G. CDRH predicts tomorrow's top technologies. *Med Dev Diag Ind* 1998; **20**: 18, 20–1.

STUDIES OF DRUG-INDUCED BIRTH DEFECTS

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.

FDA. Guidance for industry: Establishing pregnancy exposure registries. Available from <http://www.fda.gov/cder/guidance/3626fnl.pdf>.

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.

Holmes LB. Teratogen update: Bendectin. *Teratology* 1983; **27**: 277–81.

Hook EB, Healy KB. Consequences of a nationwide ban on spray adhesives alleged to be human teratogens and mutagens. *Science* 1976; **191**: 566–7.

Lenz W. Thalidomide and congenital abnormalities. *Lancet* 1962; **1**: 45.

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

Mitchell AA. Systematic identification of drugs that cause birth defects—A new opportunity. *N Engl J Med* 2003; **349**: 2556–9.

Mitchell AA. Studies of drug induced birth defects. In: Strom BL, ed. *Pharmacoeepidemiology*, 4th edn, Chichester: John Wiley & Sons, 2005; pp. 501–14.

Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. *Ann Intern Med* 1979; **90**: 257–61.

Warkany J. Problems in applying teratologic observations in animals to man. *Pediatrics* 1974; **53**: 820.

Werler MM, Mitchell AA, Hernandez-Diaz S, Honein MA. National Birth Defects Prevention Study. Use of over-the-counter medications during pregnancy. *Am J Obstet Gynecol* 2005; **193**: 771–7.

PHARMACOEPIDEMIOLOGY AND RISK MANAGEMENT

Andrews E, Gilson A, Cook S. Therapeutic risk management interventions: feasibility and effectiveness. *J Am Pharm Assoc* 2004; **44**: 491–500.

Graham DJ, Drinkard CR, Shatin D, Tsong Y, Burgess M. Liver enzyme monitoring as a risk management tool: the troglitazone experience. *JAMA* 2001; **286**: 831–3.

Graham DJ, Mosholder AD, Gelperin K, Avigan M. Pharmacoeepidemiology and risk management. In: Strom BL, ed., *Pharmacoeepidemiology*, 4th edn. New York: John Wiley & Sons, 2005; pp. 515–30.

Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002; **359**: 248–52.

Piantadosi S. *Clinical Trials: A Methodologic Perspective*. New York: John Wiley & Sons, 1997; pp. 197–8, 521.

Ray WA. Population-based studies of adverse drug effects. *N Engl J Med* 2003; **349**: 1592–4.

Rogers AS, Israel E, Smith CR, Levine D, McBean AM, Valente C, Faich G. Physician knowledge, attitudes and behavior related to reporting adverse drug events. *Arch Intern Med* 1988; **148**: 1596–600.

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.

Willy ME, Manda B, Shatin D, Drinkard CR, Graham DJ. A study of compliance with FDA recommendations for pemoline (Cylert). *J Am Acad Child Adolesc Psychiatry* 2002; **41**: 785–90.

THE USE OF PHARMACOEPIDEMIOLOGY TO STUDY MEDICATION ERRORS

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.

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. *JAMA* 1995; **274**: 29–34.

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**: 1311–16.

Berwick DM. Continuous improvement as an ideal in health care. *N Engl J Med* 1989; **320**: 53–6.

Chertow GM, Lee J, Kuperman GJ, Burdick E, Horsky J, Seger DL *et al.* Guided medication dosing for inpatients with renal insufficiency. *JAMA* 2001; **286**: 2839–44.

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

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.

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.

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.

Hazlet TK, Lee TA, Hansten PD, Horn JR. Performance of community pharmacy drug interaction software. *J Am Pharm Assoc* 2001; **41**: 200–4.

Hennessy S, Bilker WB, Zhou L, Weber AL, Brensinger C, Wang Y *et al.* Retrospective drug utilization review, prescribing errors, and clinical outcomes. *JAMA* 2003; **290**: 1494–9.

- Institute of Medicine. Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err is Human. Building a Safer Health System*. Washington, DC: National Academy Press, 1999.
- Kaushal R, Bates DW, Landrigan C, McKenna KJ, Clapp MD, Federico F, Goldmann DA. Medication errors and adverse drug events in pediatric inpatients. *JAMA* 2001; **285**: 2114–20.
- Kuperman GJ, Gandhi TK, Bates DW. Effective drug-allergy checking: methodological and operational issues. *J Biomed Inform* 2003; **36**: 70–9.
- Leape LL. Error in medicine. *JAMA* 1994; **272**: 1851–7.
- Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T *et al*. Systems analysis of adverse drug events. ADE Prevention Study Group. *JAMA* 1995; **274**: 35–43.
- Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T *et al*. Systems analysis of adverse drug events. *JAMA* 1996; **274**: 35–43.
- Leape LL, Berwick DM, Bates DW. What practices will most improve safety? Evidence-based medicine meets patient safety. *JAMA* 2002; **288**: 501–7.
- Peterson JF, Bates DW. Preventable medication errors: identifying and eliminating serious drug interactions. *J Am Pharm Assoc* 2001; **41**: 159–60.
- Schiff GD, Klass D, Peterson J, Shah G, Bates DW. Linking laboratory and pharmacy: opportunities for reducing errors and improving care. *Arch Intern Med* 2003; **163**: 893–900.
- drug reactions in two Swiss departments of internal medicine. *Br J Clin Pharmacol* 2000; **49**: 158–67.
- Finney DJ. The design and logic of a monitor of drug use. *J Chronic Dis* 1965; **18**: 77–98.
- Jha AK, Kuperman GJ, Teich JM, Leape L, Shea B, Rittenberg E *et al*. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *J Am Med Inform Assoc* 1998; **5**: 305–14.
- Joint Commission sets agenda for change. *JCAH Perspect* 1986; **6**: 6–8.
- Jones JK, Staffa J. Estimation of the frequency of warfarin-associated necrosis in a large in-patient record-linked database. *Pharmacoepidemiol Drug Saf* 1993; **2**: 115–26.
- Lamoreaux, J. The organizational structure for medical information management in the department of veterans affairs: an overview of major health care databases. *Med Care* 1996; **34**: 31–44.
- McDonald CJ, Tierney WM, Overhage JM, Martin DK, Wilson GA. The Regenstrief Medical Record System: 20 years of experience in hospitals, clinics, and neighborhood health centers. *MD Comput* 1992; **9**: 206–17.
- Onder G, Landi F, Cesari M, Gambassi G, Carbonin P, Bernabei R. Investigators of the GIFA Study. 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.
- Seidl LG, Thornton GF, Smith JW, Cluff LE. Studies on the epidemiology of adverse drug reactions. *Bull Johns Hopkins Hosp* 1966; **119**: 299–315.
- Strom BL, Gibson GA. A systematic integrated approach to improvement of drug prescribing in an acute care hospital: a potential model for applied hospital pharmacoepidemiology. *Clin Pharmacol Ther* 1993; **54**: 126–33.
- Teich JM, Merchia PR, Schmitz JL, Kuperman GJ, Spurr CD, Bates DW. Effects of computerized physician order entry on prescribing practices. *Arch Intern Med* 2000; **160**: 2741–7.

HOSPITAL PHARMACOEPIDEMIOLOGY

- Bates DW, Kuperman GJ, Rittenberg E, Teich JM, Fiskio J, Ma'luf N *et al*. A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests. *Am J Med* 1999; **106**: 261–2.
- Borda IT, Slone D, Jick H. Assessment of adverse reactions within a drug surveillance program. *JAMA* 1968; **205**: 645–7.
- Fattinger K, Roos M, Vergeres P, Holenstein C, Kind B, Masche U, Stocker DN *et al*. Epidemiology of drug exposure and adverse

The Future of Pharmacoepidemiology

The following individuals contributed to editing sections of this chapter:

BRIAN L. STROM and STEPHEN E. KIMMEL

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 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, as this book goes to press, international attention on drug safety has been higher than at any time in recent memory, as the safety of one analgesic after another is thrown into doubt, and with it the effectiveness of our entire system of drug approval and drug safety monitoring.

As the functions of academia, industry, and government become increasingly global, so does 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 700 in 2005. The International Society for Pharmacoepidemiology (ISPE), only two decades old, has grown to over 800 members from 49 countries. It has developed a set,

of guidelines for Good Epidemiologic Practices for Drug Device, and Vaccine Research in the United States in 1996, and updated these guidelines in 2004. Many national pharmacoepidemiology societies have been formed as well. The journal *Clinical Pharmacology and Therapeutics*, the major US academic clinical pharmacology journal, actively solicits pharmacoepidemiology manuscripts, as does the *Journal of Clinical Epidemiology*. The major journal of the field, *Pharmacoepidemiology and Drug Safety*, the ISPE's official journal, is now indexed on Medline. The number of individuals seeking to enter the field is rapidly increasing, 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 two decades ago the single summer short course in pharmacoepidemiology at the University of Minnesota was sometimes cancelled because of insufficient interest, more recently the University of Michigan School of Public Health summer course in pharmacoepidemiology attracted 10% of all students in the entire summer program, and now McGill University, Erasmus University Rotterdam, 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 the ISPE itself.

Regulatory bodies around the world have expanded their internal pharmacoepidemiology programs. The number of pharmaceutical companies forming their own pharmacoepidemiology units has also increased, along with their support for academic units and their funding of external pharmacoepidemiology studies. Requirements that a drug be shown to be cost effective (see Chapter 22) 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. Drug utilization review is being widely applied (see Chapter 27), and many hospitals are becoming mini-pharmacoepidemiology practice and research laboratories (see Chapter 27).

Thus, from the perspective of those in the field, the future of pharmacoepidemiology looks remarkably bright, although many important challenges remain. In this chapter, we will briefly give our own views on the future of pharmacoepidemiology. Following the format of Chapter 6 of the book, we explore this future from the perspectives of academia, the pharmaceutical industry, and regulatory agencies.

THE VIEW FROM ACADEMIA

SCIENTIFIC DEVELOPMENTS

Methodologic Advances

Methodologically, the array of approaches available for performing pharmacoepidemiology studies will continue to grow. Each of the methodologic issues discussed in Section III can be expected to be the subject of more 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 16 and 26). The next few years will likely see expanded use of neural networks, propensity scores, 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 20), especially given the recent controversies resulting from apparent inconsistencies between nonexperimental studies versus experimental studies, in some cases seeing inconsistencies where they would not really exist, if one took into

account the selection bias identified by the authors of the nonexperimental studies.

The field of pharmacoepidemiology has enthusiastically embraced the concept of therapeutic risk management (see Chapter 27). 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. Studies (i.e., program evaluations) evaluating the effectiveness of risk management programs are also badly needed. Development of this area will require considerable effort from pharmacoepidemiologists as well as those from other academic fields.

We will probably see developments in the processes used to assess causality from individual case reports (see Chapters 7, 8, and 17). We are also likely to see new guidelines emerge for the publication of spontaneous reports. "Data mining" approaches will be used increasingly in spontaneous reporting databases to search for early signals of adverse reactions (see Chapters 7 and 8). Hopefully, we will see data 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.

We are likely to see increasing input from pharmacoepidemiologists into policy questions about drug approval. 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. 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.

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 genetic factors that predispose patients to adverse drug reactions (see Chapter 18). However, few of these discoveries have yet been shown useful in improving patient care. Pharmacogenetics has evolved from studies of measures of slow drug metabolism as a contributor to adverse reactions to more recent molecular genetic markers. This has been aided by the development of new, noninvasive methods to collect DNA, such as buccal swabs, making population-based genetic studies feasible. We believe that clinical measurement of genetic 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 following up on spontaneous reports of 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 Section II of this book.

Advances can also be expected in the measurement of drug exposures in human tissue. Blood and urine have long been available for this purpose, but their collection in large ambulatory populations is difficult, and detection is largely limited to the interval shortly following exposure. For case-control studies in particular, where the exposure is measured some time after it actually occurred, such sampling is of little or no use. In recent years, however, researchers have explored the usefulness of measuring drug levels in samples of other tissues, such as hair, in which drugs or their metabolites may persist and accumulate.

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 to grow and become more innovative (see Chapter 27). Particularly as the health care industry becomes more sensitive to the possibility of overutilization, underutilization, and inappropriate utilization of drugs, and the risks associated with each, one would expect to see an increased frequency of and sophistication in drug utilization review programs, which seek to improve care (see Chapter 27). 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 our currently available drugs.

The US Joint Commission on Accreditation of Healthcare Organizations revolutionized US hospital pharmacoepidemiology through its standards requiring adverse reaction monitoring and drug use evaluation programs in every hospital (see Chapter 27). Hospitals are also now experimenting with different methods of organizing their drug delivery systems to improve their use of drugs, e.g., use of computerized physician order entry and the addition of pharmacists to ward teams (see Chapter 27).

Interest in the field of “pharmacoeconomics,” i.e., the application of the principles of health economics to the study

of drug effects, is continuing to explode (see Chapter 22). 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 requiring economic justification before permitting reimbursement for a drug. As a result, the number of pharmacoeconomic studies is increasing dramatically. As the methods of pharmacoeconomics become increasingly sophisticated, and its applications clear, this could be expected to continue to be a popular field of inquiry.

More nonexperimental studies of beneficial drug effects, particularly of drug effectiveness, can be expected, as the field becomes more aware that such studies are possible (see Chapter 21). This is being encouraged by the rapid increase in the use of propensity scores to adjust for confounding by indication, although investigators using this method often place more confidence in that technique than is warranted, 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 Chapters 16 and 21).

We will also see more use of pharmacoepidemiology approaches prior to drug approval, e.g., 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 6).

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 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 disasters associated with their use, and we hope that society and industry will turn to pharmacoepidemiologists to help evaluate the use and effects of these products.

Research interest in the entire topic of patient noncompliance (nonadherence) with prescribed drug regimens goes back to about 1960, but little fruitful research could be done until about just a decade ago because the 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 two weeks. 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. 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 judge the extent of compliance with the prescribed drug regimen. Future years are likely to see a dramatic increase in the use of this technique (see Chapter 25) 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 techniques from laboratory sciences, as described above. Statistical approaches will allow us to use predictive modeling to study, from a population perspective, who is statistically most likely to derive benefit from a drug, and who is at greatest risk of an adverse outcome. Laboratory science will enable us to determine individuals' genotypes in order to predict responses to drug therapy (i.e., molecular susceptibility) (see Chapter 18). From the perspective of pre-approval 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 genetic factors in the individualization of choice of regimens.

The past few years have seen the increased use of intermediate 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 QTc prolongation on the electrocardiogram, as a marker of risk of suffering the arrhythmia torsade de pointes, which can lead to death. Indeed, some drugs have been removed from the market, or from development, because of the presence of these intermediate markers. Yet, the utility of these markers as predictors of serious clinical outcomes is poorly studied. The next few years are likely to see data emerge to address some of these important issues.

In addition, with the growth of concerns about patient safety (see Chapter 27), there has been increasing attention to the simultaneous use of two drugs which have been shown in pharmacokinetic studies (see Chapter 4) to cause increased or decreased drug levels. Yet, there are few data indicating which, if any, of these potential interactions are of clinical importance. The next few years are likely to see the emergence of data beginning to answer this question.

Finally, in the past 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 and the smallpox vaccine program. The Institute of Medicine is currently considering the entire structure of drug development and how that should be changed, and the role that pharmacoepidemiology should play in drug development 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 (see Section II), some data resources will disappear, and a number of new computerized databases will undoubtedly emerge as major resources for pharmacoepidemiologic research, e.g., the databases from Denmark and Ontario. The importance of these databases to pharmacoepidemiology is now clear: they enable researchers to address, quickly and relatively inexpensively, questions about drug effects that require large sample sizes, with excellent quality data on drug exposures, although the data on outcomes are less certain.

Nevertheless, even as the field increases its use of databases, it is important to keep in mind the importance of studies that collect their data *de novo*. Each approach to pharmacoepidemiology has its advantages and its disadvantages, as described in Chapters 2 and 14. No approach is ideal, 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. Preferably other, 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 implements a drug benefit as part of Medicare, its health program for the elderly, government drug expenditures will suddenly be incremented by over \$40 billion annually. The Medicare drug benefit is already generating a huge new interest in pharmacoepidemiology and, if structured correctly, should generate an enormous new data resource that potentially could be useful for pharmacoepidemiologic research. 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 health care resources on drugs, yet these drugs are often used inappropriately. There have been a number of initiatives in response to this, including the World Health Organization's development of its list of "Essential Drugs."

FUNDING

For a number of years, academic pharmacoepidemiology suffered from limited research funding opportunities. With the increasing interest in the field, this situation appears to be changing, at least in the US. Much more industry funding is available, as the perceived need for the field within industry grows (see below). This is likely to increase, especially as the FDA expands its own pharmacoepidemiology program, and more often requires industry to perform postmarketing studies. This will be particularly true if these new postmarketing studies are used to permit earlier drug marketing, as has been proposed for drugs used to treat life-threatening illnesses, and has been implemented in selected situations, most notably zidovudine.

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 Healthcare Research and Quality (AHRQ) has begun to fund pharmacoepidemiologic research as well, as part of an initiative in pharmaceutical outcomes research. In particular, the AHRQ Centers for Education and Research on Therapeutics (CERTs) program appears particularly promising, to begin to provide Federal support for ongoing pharmacoepidemiology 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, and is likely to expand in the next few years.

Even the National Institutes of Health (NIH) has begun to fund pharmacoepidemiology projects more often. NIH is the logical major US source for such support, as it is the major funding source for most US biomedical research. 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 pharmacoepidemiology 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 provided most of the US government support for pharmacoepidemiology research.

It remains 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 pharmacology research. However, over the past few years it has not funded epidemiologic research, as it has focused much of its funding on molecular biology. This remains a problem for the field of pharmacoepidemiology that badly needs to be addressed. In the meantime, NIH funding is now available if one tailors a project to fit an organ system or, in some other way, to fit the priorities of one of the individual institutes.

As the US government begins to pay for drugs as part of Medicare, and therefore becomes concerned about the use, effects, and costs of drugs, it is likely that there will be substantial new funding for pharmacoepidemiology available. Indeed, as a result of the interest generated by the new Medicare drug benefit, the AHRQ has funded a new research network: *Developing Evidence to Inform Decisions About Effectiveness: The DEcIDE Network*.

Finally, but of critical importance, there is increasing concern about confidentiality in many countries. The regulatory framework for human research is actively changing, in the process. As discussed in Chapter 19, this has made 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, 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 have been 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 experienced investigators. Unfortunately, it is no longer an active program. The Merck

Foundation had a similar award for a short time, and Pfizer now funds the Pfizer Scholars Program in Clinical Epidemiology, although this has rarely gone to support investigators in pharmacoepidemiology.

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, the National Institute of General Medical Sciences (NIGMS), which funds training programs in clinical pharmacology, has not recently supported pharmacoepidemiology. This results in the dependence of pharmacoepidemiology training on non-Federal sources of funds. There are a few 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 Research Curriculum 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 is being addressed, primarily through the leadership and generosity of some pharmaceutical companies. Modest funding is also available through the CERTs program. Finally, the NIGMS is about to fund its first training program in pharmacoepidemiology, albeit a very small one, at the University of Pennsylvania School of Medicine. Much more is needed, however.

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, obviously relates to industry as well. The necessity of pharmacoepidemiology for industry has become apparent to many of those in industry (see Chapters 5 and 6). 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 pharmacoepidemiology studies “prophylactically,” to have 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 21) 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, pharmacoepidemiology 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.

In light of these advantages, most major pharmaceutical firms have formed their own pharmacoepidemiology 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 are also increasing their investment in external pharmacoepidemiologic data resources, so that they will be available for research when crises arise. All of this is likely to continue. A risk of the growth in the number of pharmacoepidemiology studies for industry is the generation of an increased number of false signals about harmful drug effects. The answer to this problem is not to avoid such studies. On the contrary, the answer is to have adequately trained individuals in the field, methodological advances within the science, and appropriate 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 6). 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, described throughout this book. Many of these crises resulted in the removal of the drugs from the market. The need for and importance of pharmacoepidemiology studies have become clear. Again, this can be expected to continue in the future. It has even been suggested that postmarketing pharmacoepidemiology studies might replace some premarketing Phase III studies in selected situations, as was done with zidovudine. 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 is increasing debate about the wisdom of developing an independent new Center for Drug Surveillance.

As noted above, the use of therapeutic risk management approaches (see Chapter 27) has been aggressively embraced by regulatory bodies around the world. This will continue to change regulation as more experience with it is gained.

Finally, as this book goes to press, there is an enormous increase in attention to drug safety, driven by drug safety issues identified with COX-2 inhibitors and even traditional NSAIDs. The net result is likely to be major regulatory changes, and possibly even new legislation, either now or in response to a new Institute of Medicine study that is nearing completion as this book goes to press.

CONCLUSION

“The person who takes medicine must recover twice, once from the disease and once from the medicine.”

William Osler, M.D.

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. 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 “thalidomide 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.

Key Points

- The discipline of pharmacoepidemiology has been growing and will likely continue to grow within academia, industry, and government.
- Methodologic advances are expected to continue in order to support pharmacoepidemiology studies, as well as newer approaches such as risk management programs and molecular pharmacoepidemiology.
- Content areas such as drug utilization review, hospital pharmacoepidemiology, pharmacoeconomics, medication adherence, patient safety, and intermediate surrogate markers will grow as interest and need for these foci increase.
- Both computerized database and *de novo* studies will continue to be important to the field and will serve as important complements to each other.
- Challenges faced by pharmacoepidemiology include limited funding opportunities, regulatory restrictions and privacy concerns surrounding human research, limited training opportunities, and inadequate personnel resources.
- All sectors responsible for the public health, including academia, industry, and government, must address the challenges facing pharmacoepidemiology and support its continued development in order to maximize the benefit and minimize the risk inherent in all medications and medical devices.

SUGGESTED FURTHER READINGS

- 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.
- Bates DW, Leape LL, Cullen DJ, Laird N, Petersen LA, Teich JM, Burdick E, Hickey M, Kleefield S, Shea B, Vander Vliet M, Seger DL. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 1998; **280**: 1311–16.
- Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. *JAMA* 1991; **266**: 2847–51.
- Howard NJ, Laing RO. Changes in the World Health Organization essential drug list. *Lancet* 1991; **338**: 743–5.
- Institute of Medicine. *The Smallpox Vaccination Program: Public Health in an Age of Terrorism*. Washington, DC: National Academies Press, 2005.
- Joellenbeck LM, Zwanziger LL, Durch JS, Strom BL, eds. *The Anthrax Vaccine: Is it Safe? Does it Work?* Washington, DC: National Academies Press, 2002.
- 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.
- Lunde PKM. WHO's programme on essential drugs. Background, implementation, present state and perspectives. *Dan Med Bull* 1984; **31** (suppl 1): 23–7.
- Pullar T, Feely M. Problems of compliance with drug treatment: new solutions? *Pharm J* 1990; **245**: 213–15.
- Spielberg SP. Idiosyncratic drug reactions: interaction of development and genetics. *Semin Perinatol* 1992; **16**: 58–62.
- Strom BL, Carson JL. Use of automated databases for pharmacoepidemiology research. *Epidemiol Rev* 1990; **12**: 87–107.
- 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.
- 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.
- Urquhart J. The electronic medication event monitor—lessons for pharmacotherapy. *Clin Pharmacokinet* 1997; **32**: 345–56.
- 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.
- Young FE. The role of the FDA in the effort against AIDS. *Public Health Rep* 1988; **103**: 242–5.
- Yudkin JS. The economics of pharmaceutical supply in Tanzania. *Int J Health Serv* 1980; **10**: 455–77.
- Yudkin JS. Use and misuse of drugs in the Third World. *Dan Med Bull* 1984; **31** (suppl 1): 11–17.

Appendix A

Sample Size Tables

Table A.1. 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	736	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 A.2. 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	1 529 057	2 153 636	4 825 616	22 279 822	28 149 090	7 764 537	2 302 889	1 183 563	755 529	540 883	415 381	278 329	147 626	99 000	41 938	15 197
0.0001	152 896	215 349	482 527	2 227 804	2 814 625	776 367	230 258	118 337	151 093	108 167	83 068	55 659	29 520	19 795	8 384	3 036
0.0005	30 570	43 057	96 475	445 402	562 673	155 196	46 024	23 651	75 539	54 077	41 528	27 825	14 756	9 895	4 189	1 516
0.001	15 280	21 521	48 218	222 602	281 179	77 550	22 994	11 815	15 095	10 805	8 297	5 558	2 946	1 974	834	300
0.005	3 047	4 292	9 613	44 362	55 984	15 433	4 571	2 346	7 540	5 396	4 143	2 774	1 469	984	414	148
0.01	1 518	2 138	4 787	22 082	27 834	7 668	2 268	1 163	1 496	1 069	820	548	288	192	79	26
0.05	295	415	927	4 258	5 315	1 456	426	216	740	528	404	269	141	93	37	11
0.10	142	200	444	2 030	2 500	680	196	97	136	95	72	47	23	14	3	—
0.15	91	128	283	1 287	1 561	421	119	58	60	41	31	19	8	3	—	—
0.20	66	92	203	916	1 092	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 A.3. 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 A.4. 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	1 285 566	1 815 876	4 068 209	18 690 665	23 293 643	6 381 472	1 869 238	950 463	601 217	427 061	325 766	215 895	112 429	74 554	30 945	11 048
0.00005	257 106	363 164	813 616	3 737 999	4 658 521	1 276 231	373 825	190 079	120 234	85 404	65 147	43 174	22 482	14 907	6 186	2 207
0.0001	128 548	181 575	406 791	1 868 916	2 329 131	638 076	186 899	95 031	60 111	42 697	32 569	21 583	11 238	7 451	3 091	1 102
0.0005	25 702	36 304	81 332	373 649	465 619	127 552	37 358	18 993	12 013	8 532	6 507	4 311	2 244	1 487	615	218
0.001	12 846	18 145	40 650	186 741	232 680	63 737	18 665	9 488	6 000	4 261	3 249	2 152	1 119	741	306	107
0.005	2 562	3 618	8 104	37 214	46 329	12 684	3 711	1 884	1 190	844	643	425	220	145	58	19
0.01	1 276	1 802	4 035	18 523	23 035	6 303	1 842	934	589	417	318	209	107	70	27	8
0.05	248	349	781	3 571	4 399	1 198	346	174	108	76	57	36	17	10	2	—
0.10	119	168	374	1 702	2 070	560	159	78	48	33	24	15	6	2	—	—
0.15	76	107	238	1 079	1 294	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 A.5. 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	78 511 105	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	15 701 142	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	7 850 222	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	1 569 25	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	784 13	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 A.6. 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	1190356	1663432	3680447	16792779	20878641	5726194	1683582	859799	546209	389547	298242	198909	104767	69986	29458	10630
0.00005	238065	332677	736066	3358436	4175543	1145183	336697	171948	109233	77903	59643	39777	20950	13994	5889	2124
0.0001	119028	166332	368018	1679143	2087555	572556	168336	85967	54611	38947	29818	19886	10473	6995	2943	1061
0.0005	23799	33257	73580	335708	417346	114455	33648	17182	10914	7783	5958	3973	2091	1396	586	210
0.001	11895	16622	36775	167779	208557	57193	16812	8584	5452	3887	2975	1983	1043	696	292	104
0.005	2372	3315	7332	33436	41526	11382	3343	1705	1082	771	589	392	205	136	56	19
0.01	1182	1651	3651	16643	20647	5656	1659	845	536	381	291	193	100	66	26	8
0.05	230	321	707	3208	3944	1075	312	157	99	69	52	34	17	10	2	—
0.10	111	154	339	1529	1856	503	144	71	44	30	23	14	6	3	—	—
0.15	71	99	216	969	1160	312	88	43	26	17	13	7	—	—	—	—
0.20	51	71	155	689	812	216	60	28	17	11	8	4	—	—	—	—
0.25	39	54	118	522	603	159	43	20	11	7	4	—	—	—	—	—
0.30	31	43	93	410	464	121	32	14	7	4	—	—	—	—	—	—
0.35	26	35	76	330	365	93	23	10	4	—	—	—	—	—	—	—
0.40	21	29	63	270	290	73	17	6	—	—	—	—	—	—	—	—
0.45	18	25	52	223	232	57	13	—	—	—	—	—	—	—	—	—
0.50	15	21	44	186	186	44	9	—	—	—	—	—	—	—	—	—
0.55	13	18	38	155	148	34	5	—	—	—	—	—	—	—	—	—
0.60	11	16	32	130	116	25	—	—	—	—	—	—	—	—	—	—
0.65	10	13	27	108	89	18	—	—	—	—	—	—	—	—	—	—
0.70	9	12	23	90	66	11	—	—	—	—	—	—	—	—	—	—
0.75	8	10	20	74	46	—	—	—	—	—	—	—	—	—	—	—
0.80	7	9	17	60	28	—	—	—	—	—	—	—	—	—	—	—
0.85	6	7	14	47	—	—	—	—	—	—	—	—	—	—	—	—
0.90	5	6	12	36	—	—	—	—	—	—	—	—	—	—	—	—
0.95	4	5	9	26	—	—	—	—	—	—	—	—	—	—	—	—

^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 A.7. 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 A.8. 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 A.9. 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	1970728	2788519	6306363	29429793	37838497	10510715	3153225	1635011	1051081	756780	583937	394159	211464	142743	61147	22330
0.00005	394143	557705	1261292	5886130	7568072	2102264	630690	327029	210236	151372	116801	78842	42300	28555	12234	4469
0.0001	197070	278853	630659	2943172	3784269	1051207	315373	163532	105130	75696	58409	39427	21155	14281	6120	2237
0.0005	39412	55772	126151	588806	757227	210362	63120	32734	21046	15155	11695	7896	4238	2862	1228	451
0.001	19704	27887	63088	294510	378847	105257	31588	16384	10535	7587	5856	3954	2124	1435	617	228
0.005	3939	5579	12638	59074	76145	21173	6363	3304	2127	1533	1184	801	432	293	128	49
0.01	1968	2790	6331	29646	38309	10663	3210	1669	1076	777	601	407	221	150	67	27
0.05	391	560	1288	6111	8059	2261	690	363	237	172	135	93	52	37	18	9
0.10	195	281	659	3181	4302	1219	379	202	133	98	77	54	32	23	13	8
0.15	129	189	451	2215	3072	879	278	150	100	75	60	43	26	19	11	8
0.20	97	143	348	1741	2476	716	230	126	85	64	52	37	23	18	11	8
0.25	77	116	287	1465	2137	624	203	113	77	59	48	35	23	18	12	9
0.30	64	98	248	1289	1930	569	188	106	73	56	46	34	23	18	13	10
0.35	56	86	222	1174	1802	536	180	103	72	56	46	35	24	19	14	11
0.40	49	77	203	1097	1727	519	177	102	72	56	47	36	25	20	15	12
0.45	44	70	191	1048	1694	513	178	104	74	58	49	38	27	22	17	14
0.50	40	66	182	1023	1696	519	182	108	77	61	52	40	29	24	19	16
0.55	38	62	178	1019	1732	535	191	114	82	66	56	44	32	27	21	18
0.60	36	61	177	1035	1806	562	203	123	89	72	61	49	36	31	25	21
0.65	35	60	180	1077	1927	605	222	135	99	80	69	56	42	36	29	25
0.70	34	61	188	1149	2110	669	248	153	113	92	79	64	49	43	35	31
0.75	35	64	203	1268	2390	764	287	178	133	109	94	77	59	52	43	38
0.80	37	70	230	1465	2831	913	348	218	164	135	117	97	75	66	55	49
0.85	43	82	278	1811	3591	1168	451	285	216	179	156	129	101	90	75	68
0.90	54	108	379	2527	5143	1687	659	420	320	266	233	195	154	137	116	105
0.95	93	190	690	4717	9851	3257	1288	828	635	531	466	391	313	280	238	217

^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 A.10. 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	1 529 065	2 153 652	4 825 672	22 280 178	28 149 758	77 647 749	2 302 966	1 183 610	755 564	540 911	415 405	278 348	147 639	99 012	41 948	15 205
0.00005	305 811	430 731	965 148	4 456 162	5 630 233	15 533 041	460 628	236 743	151 128	108 194	83 091	55 678	29 534	19 807	8 393	3 044
0.0001	152 904	215 366	482 583	2 228 160	2 815 293	7 765 578	230 335	118 385	75 573	54 105	41 552	27 844	14 770	9 906	4 199	1 524
0.0005	30 578	43 073	96 531	445 759	563 340	1 554 077	46 101	23 698	15 130	10 833	8 321	5 577	2 960	1 986	843	307
0.001	15 288	21 537	48 274	222 959	281 846	777 611	23 072	11 862	7 574	5 424	4 167	2 793	1 483	996	424	155
0.005	3 055	4 308	9 669	44 719	56 653	156 444	4 649	2 393	1 530	1 097	844	567	302	204	88	34
0.01	1 526	2 154	4 843	22 440	28 505	78 880	2 346	1 210	775	556	428	289	155	105	46	19
0.05	303	431	984	4 623	6 001	16 744	506	264	171	124	97	66	37	26	13	7
0.10	150	216	503	2 405	3 207	904	279	148	97	71	56	39	23	17	9	6
0.15	100	145	343	1 673	2 292	653	205	111	74	55	44	31	19	14	8	6
0.20	74	110	265	1 313	1 849	533	170	93	63	47	38	28	17	13	8	6
0.25	59	89	218	1 104	1 597	465	151	84	57	44	35	26	17	13	9	6
0.30	49	75	188	971	1 443	425	140	79	55	42	34	26	17	14	9	7
0.35	42	65	168	883	1 349	401	135	77	54	42	34	26	18	14	10	8
0.40	37	58	154	825	1 294	388	133	77	54	42	35	27	19	15	11	9
0.45	33	53	144	788	1 270	385	133	78	56	44	37	28	20	17	13	10
0.50	31	50	137	768	1 272	389	137	81	58	46	39	31	22	19	14	12
0.55	28	47	133	764	1 301	402	144	86	62	50	42	33	24	21	16	14
0.60	27	45	133	775	1 357	423	154	93	68	55	47	37	28	24	19	16
0.65	26	45	135	805	1 449	456	168	103	76	61	52	42	32	28	22	19
0.70	26	45	140	859	1 588	505	188	116	86	70	61	49	38	33	27	23
0.75	26	47	151	947	1 799	577	218	136	102	84	72	59	46	40	33	29
0.80	28	51	170	1 092	2 133	690	265	166	125	104	90	74	58	51	42	38
0.85	31	60	205	1 349	2 708	884	343	218	165	137	120	100	78	70	58	53
0.90	39	78	279	1 880	3 881	1 278	503	322	246	205	180	150	119	107	90	82
0.95	66	137	506	3 505	7 438	2 472	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 A.11. 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	1 369 478	1 930 861	4 322 663	19 888 975	24 913 964	6 843 813	2 014 824	1 029 056	653 448	465 720	356 295	237 271	124 583	83 040	34 800	12 517
0.00005	273 893	386 172	864 545	3 977 907	4 983 044	1 368 844	402 996	205 830	130 703	93 155	71 268	47 461	24 922	16 612	6 963	2 506
0.0001	136 945	193 086	432 280	1 989 023	2 491 679	684 473	201 517	102 927	65 360	46 584	35 640	23 735	12 464	8 309	3 483	1 254
0.0005	27 387	38 617	86 468	397 917	498 587	136 977	40 334	20 604	13 086	9 328	7 137	4 754	2 498	1 666	700	253
0.001	13 692	19 309	43 242	199 028	249 451	68 540	20 186	10 314	6 551	4 671	3 574	2 382	1 252	836	352	128
0.005	2 736	3 862	8 661	39 918	50 143	13 790	4 068	2 082	1 324	945	724	484	256	171	73	28
0.01	1 367	1 931	4 338	20 030	25 231	6 947	2 054	1 053	671	480	368	246	131	88	39	16
0.05	271	387	881	4 125	5 313	1 477	444	231	149	108	84	57	32	22	11	6
0.10	134	194	450	2 145	2 841	799	245	129	85	62	49	34	20	14	8	5
0.15	89	130	307	1 491	2 031	577	180	97	64	48	38	27	16	12	7	5
0.20	66	98	236	1 171	1 639	471	150	82	55	41	33	24	15	12	7	5
0.25	53	79	195	984	1 417	412	133	74	50	38	31	23	15	12	8	6
0.30	44	67	168	865	1 281	376	124	70	48	37	30	23	15	12	8	6
0.35	38	58	150	786	1 197	355	119	68	47	37	30	23	16	13	9	7
0.40	33	52	137	734	1 149	345	118	68	48	37	31	24	16	14	10	8
0.45	30	47	128	700	1 128	342	119	69	49	39	32	25	18	15	11	9
0.50	27	44	122	682	1 131	346	122	72	52	41	35	27	19	16	12	10
0.55	25	42	119	679	1 156	357	128	76	55	44	38	30	22	18	14	12
0.60	24	40	118	689	1 207	377	137	83	60	49	41	33	25	21	17	14
0.65	23	40	119	715	1 289	406	150	91	67	55	47	38	28	24	20	17
0.70	23	40	124	762	1 414	450	168	104	77	63	54	44	33	29	24	21
0.75	23	42	133	839	1 602	515	195	121	91	75	65	53	41	36	29	26
0.80	24	45	150	968	1 900	616	236	149	112	93	80	66	52	45	38	34
0.85	27	52	180	1 194	2 413	789	307	195	148	123	107	89	70	62	52	46
0.90	34	68	245	1 664	3 459	1 142	450	288	220	184	161	134	107	95	80	72
0.95	57	119	444	3 100	6 632	2 208	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 A.12. 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 A.13. 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	1 472 099	2 082 974	4 710 741	21 983 531	28 264 683	7 851 317	2 355 404	1 221 324	785 139	565 302	436 191	294 430	157 960	106 627	45 676	16 681
0.00005	294 418	416 596	942 163	4 396 835	5 653 216	1 570 354	471 115	244 286	157 043	113 073	87 248	58 894	31 598	21 330	9 139	3 339
0.0001	147 208	208 299	471 091	2 198 497	2 826 782	785 234	235 579	122 156	78 531	56 544	43 631	29 452	15 803	10 668	4 572	1 671
0.0005	29 440	41 661	94 233	439 828	565 636	157 137	47 150	24 452	15 721	11 321	8 736	5 899	3 166	2 138	918	337
0.001	14 719	20 831	47 126	219 994	282 992	78 625	23 596	12 239	7 870	5 668	4 375	2 954	1 587	1 072	461	171
0.005	2 943	4 168	9 441	44 128	56 879	15 816	4 753	2 469	1 589	1 146	885	599	323	219	96	37
0.01	1 470	2 085	4 730	22 145	28 617	7 966	2 398	1 248	804	581	449	305	165	113	50	20
0.05	293	419	962	4 566	6 020	1 690	516	272	177	129	101	70	39	28	14	7
0.10	146	211	493	2 377	3 214	911	283	151	100	74	58	41	24	18	10	6
0.15	97	142	337	1 655	2 295	657	208	113	75	56	45	32	20	15	9	6
0.20	73	107	260	1 301	1 850	535	172	95	64	48	39	28	18	14	9	6
0.25	58	87	215	1 095	1 597	466	152	85	58	44	36	27	17	14	9	7
0.30	49	74	186	964	1 442	425	141	80	55	42	35	26	18	14	10	8
0.35	42	65	166	877	1 346	401	135	77	54	42	35	26	18	15	11	9
0.40	37	58	152	820	1 291	388	133	77	54	42	35	27	19	16	12	10
0.45	33	53	143	784	1 266	384	133	78	56	44	37	29	20	17	13	11
0.50	31	50	137	765	1 267	388	137	81	58	46	39	31	22	19	15	12
0.55	29	47	133	761	1 294	400	143	85	62	50	42	33	25	21	16	14
0.60	27	46	133	774	1 350	421	152	92	67	54	46	37	28	24	19	16
0.65	26	45	135	805	1 440	453	166	101	75	61	52	42	32	27	22	19
0.70	26	46	141	859	1 577	500	186	115	85	69	60	49	37	32	26	23
0.75	27	48	152	948	1 785	571	215	134	100	82	71	58	45	39	33	29
0.80	28	53	172	1 095	2 115	682	260	163	123	101	88	73	57	50	42	37
0.85	32	62	208	1 353	2 683	873	337	213	162	134	117	97	76	68	57	51
0.90	41	81	283	1 888	3 842	1 260	493	314	240	200	175	146	116	103	87	79
0.95	70	142	516	3 524	7 359	2 433	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 A.14. 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	1190363	1663444	3680489	16793046	20879138	5726351	1683639	859834	546235	389568	298260	198923	104777	69995	29465	10635
0.00005	238071	332689	736108	3358703	4176039	1145339	336754	171983	109259	77923	59660	39791	20960	14003	5896	2129
0.0001	119034	166344	368060	1679410	2088152	572713	168393	86001	54637	38967	29835	19899	10483	7004	2950	1066
0.0005	23805	33269	73622	335976	417842	114612	33705	17216	10939	7803	5975	3986	2101	1405	593	216
0.001	11901	16635	36817	168047	209054	57349	16869	8618	5477	3907	2993	1997	1053	705	298	109
0.005	2378	3327	7374	33704	42024	11540	3400	1740	1107	791	607	406	215	145	63	24
0.01	1188	1664	3693	16911	21146	5814	1717	880	561	402	308	207	211	75	33	14
0.05	236	333	750	3482	4455	1237	371	193	125	91	70	48	27	19	10	5
0.10	117	167	383	1810	2383	669	205	109	71	53	41	29	17	13	7	5
0.15	77	112	261	1258	1704	484	152	82	54	41	32	23	14	11	7	5
0.20	58	84	201	987	1376	396	126	69	47	35	28	21	13	10	7	5
0.25	46	68	166	829	1190	346	112	62	43	33	27	20	13	10	7	5
0.30	38	57	143	729	1076	316	105	59	41	32	26	20	13	11	7	6
0.35	33	50	127	662	1006	299	101	58	40	31	26	20	14	11	8	7
0.40	29	45	116	618	966	290	99	58	41	32	27	21	15	12	9	7
0.45	26	41	108	590	949	288	100	59	42	33	28	22	16	13	10	8
0.50	24	38	103	574	951	292	103	61	44	35	30	24	17	15	11	10
0.55	22	36	100	571	973	301	108	65	47	38	32	26	19	16	13	11
0.60	21	34	99	579	1016	318	116	70	52	42	36	29	22	19	15	13
0.65	20	34	101	601	1085	343	127	78	58	47	40	33	25	22	18	16
0.70	20	34	105	640	1190	380	143	89	66	54	47	38	29	26	21	19
0.75	20	35	112	705	1350	435	166	104	78	64	56	46	36	32	26	23
0.80	21	38	126	812	1601	520	201	127	96	80	70	58	45	40	34	30
0.85	23	44	152	1002	2034	667	261	167	127	106	93	77	61	55	46	42
0.90	29	58	205	1395	2916	965	383	246	189	158	139	117	94	84	71	65
0.95	48	100	371	2598	5592	1868	750	487	376	316	279	236	190	171	147	134

^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 A.15. 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 A.16. 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	1 034 611	1 440 327	3 154 151	14 188 340	17 179 015	4 657 221	1 342 151	674 222	422 474	297 830	225 778	148 193	76 026	49 982	20 443	7 227
0.00005	206 920	288 065	630 838	2 837 745	3 435 981	931 503	268 452	134 858	84 505	59 574	45 162	29 644	15 209	9 999	4 091	1 447
0.0001	103 459	144 032	315 424	1 418 920	1 718 102	465 788	134 240	67 438	42 259	29 792	22 585	14 825	7 607	5 002	2 047	725
0.0005	20 690	28 806	63 092	283 861	343 799	93 216	26 870	13 501	8 462	5 966	4 524	2 970	1 525	1 003	412	147
0.001	10 344	14 403	31 551	141 978	172 011	46 645	13 449	6 759	4 237	2 988	2 266	1 489	765	504	207	75
0.005	2 067	2 880	6 318	28 473	34 581	9 388	2 712	1 366	858	606	460	303	157	104	44	17
0.01	1 032	1 440	3 164	14 285	17 404	4 731	1 370	691	435	308	234	155	81	54	23	10
0.05	205	288	641	2 938	3 670	1 009	298	153	98	70	54	37	20	14	7	4
0.10	101	144	327	1 525	1 966	547	166	87	57	41	32	23	13	10	5	3
0.15	67	96	222	1 059	1 408	397	123	66	43	32	26	18	11	8	5	4
0.20	50	72	171	830	1 138	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	6	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	1 131	366	140	88	67	55	48	39	31	27	22	20
0.80	17	31	103	670	1 343	439	171	108	82	68	60	50	39	35	29	26
0.85	18	35	123	826	1 708	564	222	143	109	91	80	67	53	47	40	36
0.90	23	45	166	1 148	2 452	817	327	211	163	137	120	101	81	73	62	56
0.95	37	78	298	2 133	4 707	1 583	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 A.17. 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.

An *adverse drug event* or *adverse drug experience* is an untoward outcome that occurs during or following clinical use of a drug, whether preventable or not.

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.

Studies of *adverse events* explore any medical events experienced by patients and use epidemiologic methods to investigate whether any given event occurs more often in those who receive a drug than in those who do not receive the drug.

An *adverse experience* is any adverse event associated with the use of a drug or biological product in humans, whether or not considered product related.

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, typically 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 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.

Autocorrelation is where any individual observation is to some extent a function of the previous observation.

Bias is a systematic manner in which the two study groups have been treated differently. The presence of a bias causes a study to yield incorrect results.

Biological inference is the process of generalizing from a statement about an association seen in a population to a causal statement about biological theory.

Case-cohort studies are studies that compare cases with a disease to a sample of predetermined size of subjects randomly selected from the parent cohort, analyzing the study as a cohort study.

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 with a disease to a different time period in the same individuals, looking for differences in antecedent exposures.

Case reports are reports of the experience of single patients. As used in pharmacoepidemiology, a case report describes a single patient who was exposed to a drug and experiences 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.

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.

Drug clearance is the proportion of the “apparent” volume of distribution that is cleared of drug over a specified time. The total body clearance is the sum of clearances by different routes, e.g., renal, hepatic, pulmonary, etc.

Clinical pharmacology is the study of the effects of drugs in humans.

Cohort studies are studies that identify defined populations and follow them forward in time, examining their rates of disease. Cohort studies generally identify and compare exposed patients to unexposed patients or to patients who receive a different exposure.

Confidence interval is a range of values within which the true population value probably lies.

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 create the appearance of an association between the risk factor and the outcome or mask a real one.

Confounding by indication can occur when the underlying diagnosis or other clinical features that trigger use of a certain drug also are related to patient outcome.

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.

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 determined 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 between two or more interventions.

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 gold standard.

Cross-sectional studies examine exposures and outcomes in populations at one point in time; they have no time sense.

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 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 nonmedical costs are nonmedical care costs incurred because of an illness or the need to seek medical care. They can include the cost of transportation to the hospital or physician's office, the cost of special clothing needed because of the illness, and the cost of hotel stays and special housing (e.g., modification of the home to accommodate the ill individual).

Discriminative instruments are those that measure differences among people at a single point in time.

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). 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 diseases 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 patients 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.

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. $T_{1/2}$ is a function of both the volume of distribution and clearance of the drug.

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 the cost of morbidity (e.g., time lost from work) or 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 (often 10 or more) of study populations before and after an intervention.

Medication errors are defined as any mistake in the medication use process, including prescribing, transcribing, dispensing, administering, and monitoring.

Meta-analysis is the 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.

Near misses are medication errors that have high potential for causing harm but did not, either because they were intercepted prior to reaching a patient or because the error reached the patient who fortuitously did not have any observable untoward sequelae.

Nondifferential misclassification occurs when the misclassification of one variable does not vary by the level of another variable. Nondifferential misclassification usually results in bias toward the null.

Nonexperimental studies are studies in which the investigator does not control the therapy, but observes and evaluates the results of ongoing medical care. These are the study designs that do not involve random allocation, namely case reports, case series, analyses of secular trends, case-control studies, and cohort studies.

Observational studies are studies in which the investigator does not control the therapy, but observes and evaluates the results of ongoing medical care. These are the study designs that do not involve randomization, namely 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 nondiseased group. It provides an estimate of the relative risk when the disease under study is relatively rare.

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.

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.

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

Pharmacogenetic epidemiology is the study of the effects of genetic determinants of drug response on outcomes in large numbers of people.

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 of a drug administered and the serum or blood level achieved. It includes the study of the processes of drug absorption, distribution, metabolism, and excretion.

Pharmacology is the study of the effects of drugs in a living system.

Pharmacotherapeutics is the application of the principles of clinical pharmacology to rational prescribing, the conduct of clinical trials, and the assessment of outcomes during real-life clinical practice.

Pharmacovigilance is the scientific and data gathering activities relating to the detection, assessment, and understanding of adverse event reports.

Pharmionics is the study of how patients use or misuse prescription drugs in ambulatory care.

Postmarketing surveillance is the study of drug use and drug effects after release onto the market. This term is some-

times used synonymously with “pharmacoepidemiology,” but the latter can be relevant to premarketing studies as well. Conversely, the term “postmarketing surveillance” is sometimes felt to apply 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.

Potential adverse drug events are medication errors that have high potential for causing harm but did not, either because they were intercepted prior to reaching a patient or because the error reached the patient who fortuitously did not have any observable untoward sequelae.

The *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 nonrandomly selected group exposed to a treatment (e.g., physicians receiving feedback on specific prescribing practices), as well as simultaneous before and after observations of a similar (comparison) group not receiving treatment.

Prescribing errors refer to issues related to underuse, overuse, and misuse of prescribed drugs, all of which contribute to the suboptimal utilization of pharmaceutical therapies.

The *prevalence* of a disease is a measurement of how common the disease is. Specifically, it is the number of existing cases of the disease in a defined population at a given point in time or over a defined time period, divided by the number of people in that population.

Prevalence study bias is a type of selection bias that may occur in studies when prevalent cases rather than new cases of a condition are selected for a study. A strong association with prevalence may be related to the duration of the disease rather than to its incidence, because prevalence is proportional to both incidence and duration of the disease.

Privacy, in the setting of research, refers to each individual's right to be free from unwanted inspection of, or access to, personal information by unauthorized persons.

Propensity scores are an approach to controlling for confounding that uses mathematical modeling to predict

exposure, rather than the traditional approach of predicting *outcome*.

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.

Publication bias occurs when publication of a study’s results is related to the study’s findings.

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, e.g., when the use of the drug contributes to the diagnostic process.

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 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.

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)].

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 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.

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.

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 error introduced into a study by its design, rather than 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, dose-related, and less serious than Type B reactions.

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 error* is concluding that there is an association when in fact one does not exist, i.e., erroneously rejecting the null hypothesis.

A *type II error* is concluding that there is no association when in fact one does exist, i.e., erroneously accepting the null hypothesis.

Unanticipated beneficial effects of drugs are desirable effects that could not have been predicted on the basis of existing knowledge.

Unanticipated harmful effects of drugs are unwanted effects that could not have been predicted on the basis of existing knowledge.

Uncontrolled studies refer to studies without a comparison group.

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, e.g., 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

Note: Page numbers in *italics* refer to figures and tables

- absorption of drug 42–3
- academic detailing programs 406
- academic institutions
 - pharmaceutical industry
 - relationship 75
 - research funding 451
- academic integrity 68
- academic medical centers (AMCs)
 - 63, 65
 - care delivery models 67
 - medical therapeutics 65–6
 - practice improvement 65
 - researcher training/support 67
 - role 65–8
- Accident and Emergency
 - department data 200
- accountability in regulatory action 85
- acetaminophen 48, 142
- acetazolamide 132
- acetylation polymorphism 46–7
- N*-Acetylbenzo-quinoneimine 48
- acne 206–7
- adherence, prescribers to treatment
 - guidelines 404
- adherence to treatment 247, 367–81, 447–8
 - antihypertensive drugs 378
 - clinical problems 367–9
 - electronic medication event
 - monitoring 370
 - estimates 369–70
 - ethics 374
 - medication errors 427
 - methodologic problems 369–70
 - once-daily dosing 373–4
 - terminology 368–9
 - time course 367, 369
 - twice-daily dosing 376–7
 - variations 367–8, 375, 377
- administration route errors 429
- administrative databases 253–4
- adrenergic receptors, cell surface 38
- β_2 -adrenergic receptors 289–90
- adverse drug events (ADEs) 477
 - aggregate analysis tools 109
 - bias 111
 - causal relationship 278
 - community setting 431–2
 - data mining 109
 - drug holidays 378
 - incidence 10
 - density 160
 - medication errors 427
 - near misses 428
 - NSAIDs 311–12
 - potential 425, 476
 - prescriber behavior 50
 - Prescription-Event Monitoring 157
 - rare 384
 - serious 155
 - rates 160
 - reappearance 131
 - recognition 111
 - registry 5
 - regulatory action 107
 - safety of drugs 78
 - studies 4
 - time 186
- vaccines 411
 - see also* reports/reporting
- adverse drug reactions (ADRs) 469
 - case reports 57
 - causality 278–80
 - delayed 60
 - denominator for calculations 434
 - FDA surveillance program 93
 - frequency 83
 - hospital-based monitoring 434–5
 - immediate 60
 - incidence 3, 18
 - long latency 161–2
 - mechanisms 132–3
 - monitoring 447
 - population size 18
 - post-licensing 80
 - profile 83
 - rare 33
 - serious events 78, 101–2
 - signal detection 80–2
 - suspected unexpected serious 78
 - systematic manual review 81
 - time of occurrence 432
 - tolerability 59–60
 - Types A and B 4, 60, 479
 - see also* causality;
 - reports/reporting
- adverse effects studies 4
 - meta-analyses 357–60
- Adverse Event Reporting System (AERS) 93, 104–5
 - applications 113–14
 - database 283

- adverse experiences 94, 473
 - serious 94, 479
 - unexpected 94, 479
- age
 - confounding 266
 - effect modification 267
 - pharmacodynamic effects 39
 - see also* children; elderly people
- Agency for Healthcare Research and Quality (US) 8, 63–4, 447, 451
- aggregate analysis tools 109
- agranulocytosis 131, 132
- airflow limitation, chronic 347
- albuterol 269, 289
- algorithms of causality 282
 - requiring scoring of individual judgments 281–3
- alkylating agents 72
- allergy errors 429
- alosetron 7
- alternative medicine 449
- ambulance data 200
- American Dental Association 102
- American Hospital Formulary Society therapeutic categories 403
- American Medical Association (AMA) 93, 102
- analytic studies 22, 76, 473
- Anatomic Therapeutic Chemical (ATC) classification 403
- angina 284
- angiotensin converting enzyme (ACE) inhibitors 45, 83, 290, 335–6
- angiotensin receptor antagonists 45
- anthrax vaccine 7
- antibiotics
 - acne 206–7
 - Clostridium difficile* association 387–8
- antidepressants 177–8
- antihypertensive drugs 377
- antipsychotic drugs 191–2
- Apotek AB 217
- appetite suppressants 40–1
- aprimidine 131
- aspirin 360
 - gastric ulceration 38–9
- associations 14, 473
 - artifactual 14–15
 - causal 15
 - criteria 15–17
 - confounded 15
 - consistency 16
 - measures 246–7
 - strength 16
- types 14–15
 - unsuspected 143–4, 145
- astemizole 6
- asthma 326, 349, 384, 388, 389, 391
 - computerized decision support 405–6
- Asthma Quality-of-Life Questionnaire (AQLQ) 349
- atenolol 378
- audit 85
- authorization bias 306–7
- autobiographical memory, cognitive theory 240–2
- autonomy, respect for 305
- azapropazone 360
- Bayes probability theorem, causality assessment 280, 283–4
- Bayesian approach
 - angina and NSAIDs 284
 - meta-analysis 356
- Bayesian Confidence Propagation Neural Network (BCPNN) 70, 124–5, 128
 - data mining validation 125–6
- BCG vaccination 131–2
- Bendectin® 419, 422–3
- bendrofluthiazide 373
- benefit–risk assessment 82–3, 85
- benefits of treatment 50–1
 - action to increase 85
 - balance with risk 83
 - evaluation 82–3
- benzodiazepines 40, 46, 145
 - use reduction 410
- benzyl alcohol 36
- beta agonists 349, 383–5, 388
- beta blockers 38, 60, 289–90, 361
 - myocardial infarction 324
 - non-intrinsic sympathomimetic action 379
- bias 15, 272, 473
 - adverse drug events 111
 - ascertainment 434
 - authorization 306
 - case–control surveillance 146–7
 - cognitive 51
 - data abstraction 354–5
 - detection 265, 474
 - effect measures 244
 - hypothesized source 357
 - immortal time 384, 394
 - incidence 476
 - information 261, 262–3, 268, 427, 473
 - misclassification 262–3, 474
 - National Childhood Encephalopathy Study 261
- pharmacoepidemiology 261–71
- prevalence 262, 476, 477
- protopathic 261, 262, 478
- publication 354, 357, 478
- recall 146–7, 249, 264, 420, 478
- referral 261–2
 - inpatients 434
- reporting 130
- sample distortion 262
- selection 146, 220, 247, 262, 267–8, 474
- self-selection 262
- suspicion of clinical event 277
- vaccines 411
- bioavailability of drugs 48
- bioequivalence 330
- bioethics 301–9
 - research 302–5
- bioinformatics 448
- biological inference 14, 473
- biological plausibility 15–16
- biological terrorism 64
- Biologics Control Act (1902) 92
- biomarkers
 - interactions with medication 292–4
 - intermediate 450
- biotechnology revolution 334
- birth, drug evaluation of outcomes 73
- birth defects, drug-induced 418–23, 440
 - biologic plausibility 420
 - case–control studies 420
 - clinical problems 418
 - cohort studies 420
 - confounding 420
 - data sets for other purposes 421
 - exposure 417
 - large population studies 421
 - methodologic problems 419–20
 - outcome 418
 - recall bias 420
 - sample size 419
 - secular changes 422
 - selected exposures 421
 - statistical power 421
 - validity 422
 - see also* teratogens
- Black Box Warnings 184
- blinding 313
 - economic analysis 341
 - information bias 268
- bootstrap methods, nonparametric 342
- Boston Collaborative Drug Surveillance Program 6, 120, 434
- Bradford Hill criteria 278
- breast cancer, metastatic 334–5

- breast implants, silicone 6
- Brigham and Women's Hospital inpatient database 432–3
- Brigham Integrated Computing System (BICS) 435
- bromocriptine 6
- buccal cell samples, case–control surveillance 142
- budgets, prescribing 51
- bupropion 162
- calcium antagonists 49
- calcium channel blockers 197
- cancer
 - screening 329
 - services data 193
 - in utero drug exposure 73
 - see also named cancers*
- candidate gene association studies 292, 295
- carbamazepine 418
 - metabolites 48
- cardiac glycosides 38
- cardiac output 41
- carvedilol 164
- case–cohort design 389–90, 474
- case–control studies 19, 20–1, 22, 470
 - ad hoc* 219–20
 - appetite suppressants 40–1
 - birth defects 419
 - case–control surveillance 138
 - confounding 384
 - controls 313
 - crossover model 391–2
 - data sources 218–23
 - database resources 232
 - effect modification 293
 - effectiveness measurement 328, 330
 - exposure
 - low prevalence 235
 - measurement 448
 - gene–drug interactions 290–1
 - incidence outcome 234
 - nested 163, 387–8, 390
 - sample size 29, 30, 31, 32, 33, 459–66
- case–control surveillance (CCS) 137–48
 - applications 147
 - bias 146–7
 - buccal cell samples 142
 - case specification 142
 - confounding control 146
 - confounding variables 143
 - control specification 142
 - data analysis 142–4
 - database 138, 146
 - diagnoses 139
 - dietary supplements 141, 145
 - DNA collection 142
 - drug dictionary 142
 - drug–disease associations 141
 - drug/genotype analysis 143
 - drug information 139, 140, 141
 - drug use 142–3
 - duration of use 144
 - effect modification 143
 - effects after long intervals 145
 - genetic polymorphisms 146
 - hospital records information 141–2
 - hypothesis testing 142–3
 - interviews 138, 139, 144
 - methods 138–9, 140, 142
 - modifying factors 141
 - non-prescription medications 145
 - nondifferential misclassification 147
 - outcome data 146
 - participation rates 139
 - statistical power 143, 146
 - unsuspected associations 143–4, 145
- case–crossover designs 391–3, 474
- case definitions 219
- case identification 186
- case–non case method 123
- case reports 18, 21–2, 473
 - adverse reactions 57
 - causation 277–86, 448
 - risk management 423
 - structured approach 280
- case series 18–19, 21–2, 474
 - sample size 31–2
- case–time–control design 391, 393
- causal inference 377
- causal relationships 131
 - assessment 121–2
 - observational studies 312
- causality 83, 473
 - accrued value of evaluations 283
 - accuracy of judgment 283
 - adverse drug reactions 278–80
 - algorithms 281
 - requiring scoring of individual judgments 281–3
 - assessment 121–2
 - applications 285
 - Bayes probability theorem 279, 281
 - comparison of methods 283–5
 - methodologic problems 280–1
 - uses 279–80
- automation of evaluation 285
- case reports 277–86, 448
- criteria 278–80, 282
- criterial method with verbal judgments 281
- critical elements 285
- gene–drug interactions 290
- global introspection 280–2
- likelihood
 - assessment 82
 - determination methods 277–81
 - number of evaluations 283
 - probabilistic methods 281–3
 - proof 110
 - unstructured clinical judgment 280
 - vaccine adverse events 411
- causation 17
- celecoxib 152, 154, 160
- cell replication blocking 37
- cell surface receptors 37, 38
- Center for Drug Evaluation and Research 99–100
- Center for Food Safety and Nutrition (CFSCAN), Adverse Event Reporting System 99
- Centers for Disease Control and Prevention 453
- Centers for Education and Research on Therapeutics (CERT) 8, 63–4, 65, 182, 447
 - databases 66
 - public–private partnerships 68
- Centers for Medicare and Medicaid Services 188
- central nervous system (CNS)
 - depressants 40
- cerivastatin 7, 134
- challenge–rechallenge 18
- change, global ratings 347
- changeability 474
 - HRQL instruments 347
- channeling of drugs 392
- chart review 428
- chelating agents 43
- chemotherapy 72
 - plus autologous stem cell transplantation 334–5
- children
 - drug use 323
 - medication errors 427
- chlamydial infection 372
- chloramphenicol 93
- cholestyramine 43
- Chronic Respiratory Questionnaire 348
- cigarette smoking
 - asthma 389
 - lung cancer 15, 16, 17, 19

- cimetidine 18, 49, 56
 - liver disease 71
 - postmarketing surveillance 223
 - claims-based evaluations, medication
 - errors 427, 428
 - claims databases 168–9, 170, 176, 187
 - classification validity 246
 - clearance of drug 42, 474
 - cleft palate 421
 - clinical economics 336–8
 - clinical judgment 369
 - unstructured 282
 - clinical outcomes, efficacy of drug 82–3
 - clinical pharmacology 4, 36, 474
 - funding 66
 - clinical significance 33
 - clinical trials 69–70
 - economic analysis 338–9
 - inclusion/exclusion criteria 69–70
 - patient sample size 69
 - safety data 78–9
 - usual care arm 339
 - see also named types of trials*
 - clioquinol 6
 - Clostridium difficile* 387–8
 - clusters of events 131
 - cognitive bias 51
 - cognitive processes, temporal questions
 - 241
 - cognitive theory, autobiographical
 - memory 240–2
 - cohort effect 386
 - cohort studies 19–21, 22, 473
 - ad hoc* 222–3, 235
 - new drugs 235
 - confounding 265
 - controls 313
 - data sources 220
 - database resources 232
 - drug exposure changes 384
 - drug-induced birth defects 418–19
 - fixed entry 384
 - losses to follow-up 262
 - Phase IV 56
 - postmarketing 137–8
 - risk sets 386–7
 - sample size 26–7, 28, 29, 32, 451–9
 - sampling within 384–8
 - structures 384–6
 - study design 383–4
 - types 386–8
 - vaccine efficacy 329
 - variable entry 386
 - Collaborative Perinatal Project (CPP)
 - 421
 - colorectal cancer risk and NSAIDs
 - 263–4
 - common practice, economic analysis
 - 338–9
 - Common Rule 302–3, 304–5
 - communications
 - cross-cultural 64
 - feedback to reporters 122
 - initiatives 135
 - regional centers 120
 - risk 186
 - safety of drug 85
 - communitarianism 304
 - community setting, adverse drug events
 - 428–9
 - compassionate use 78
 - complementary medicine 449
 - compliance
 - large simple trials 317
 - white-coat 370
 - see also adherence to treatment*
 - compliance-dependent drugs 375, 378
 - computer technology 64, 256
 - Computerized Online Medicaid Analysis
 - and Surveillance System 6
 - computerized prescriber order entry
 - (CPOE) programs 406, 408, 433
 - conditional logistic regression 389
 - confidence intervals 20–1, 449, 474
 - adverse drug reactions 78–9
 - Poisson-distributed variables 471
 - confidentiality 474
 - adverse event reporting 103
 - information use 307
 - patient records 256
 - research participants 75, 77, 303–4, 305–6, 307–8, 447
 - Tayside Medicines Monitoring Unit 198, 207
 - violation risk 301–2, 304
 - confounding 36, 263–7, 272, 474
 - age 267
 - analysis level 270–1
 - angina and NSAIDs 284
 - case–control studies 383
 - cohort studies 265
 - comedication 268
 - control 233, 268–71
 - case–control surveillance 145
 - database resources 270–1
 - design level 270–1
 - drug-induced birth defects 418–19
 - effect modification 268–9
 - by ethnicity 294–5, 297
 - by indication 267–8, 273–4, 313–14, 324–5, 326, 420, 474
 - indication for prescription 267–8
 - large simple trials 314
 - matching 270
 - mechanism 264–5
 - modeling 271
 - multivariate analysis 272
 - odds ratio 391
 - partial data 383, 390–1
 - population admixture 294–5
 - propensity scores 273
 - randomization 268–9, 313
 - stratification 272–3
 - two-stage sampling 272–3, 391
 - uncertainty source 384
 - uncontrolled 312
 - vaccines 411
- confounding variables 15, 474
 - automated databases 168
 - case–control surveillance 138, 145
 - database resources 235
 - information need 194–5
 - restriction to one level 271
- congenital anomalies 162
- consent *see* informed consent
- construct validity 346–7, 474
- consumer behavior 50–2
- consumer reports 118
- contact lenses, soft 417
- continuous-benefit effect model 341
- controls/control groups 22
 - absence 222–3, 318
 - choice of treatment 313–14
 - expected incidence of outcome 26
 - friend 219–20
 - identification 186
 - number of groups 27
 - number of unexposed subjects 27
 - placebo 314
 - recruitment 219
- corneal ulcers 417
- coronary artery bypass surgery (CABG)
 - 361
- Coronary Drug Project 268
- correlation coefficients 245
- corticosteroids 326
- cost(s) 474
 - analysis methods 341–2
 - analysis plan 341
 - development of drugs 74
 - direct 337–8, 475
 - drug statistics 400
 - drug use 10–11
 - prescriber behavior 52

- identification 336
- indirect 476
- intangible 337, 338, 476
- medical care 333–4
- meta-analysis impact 361–2
- productivity 337, 338
- studies 55
- total 401
- types 337–8
- unit 401
- cost-benefit analysis 336, 474
- Cost Containment Program 436–7
- cost-effectiveness 51, 474
 - analysis 336–7
 - molecular pharmacoepidemiology 297
 - postmarketing drug surveillance 110
 - requirements 448
 - studies 328–9
- cost-efficacy analyses 339
- cost-identification analysis 474
- cost-utility analysis 350
- coumarin 46
- Council for International Organizations of Medical Sciences (CIOMS) 82
 - ethics board review 303
 - international reporting standards 128, 129
 - monograph IV 133
- counterfeit medicines 135
- COX-2 inhibitors 39, 152, 160
- creatinine
 - clearance 44, 45
 - plasma concentration 44
- Creutzfeldt-Jacob disease (CJD) 72
- critical method with verbal judgments 281
- criterion validity 346, 474
- cyclosporine 45, 49
- CYP2C9 polymorphism 291
- cytochrome P450 (CYP) 46, 48, 49
- data
 - abstraction bias 354–5
 - access to existing 307–8
 - administrative databases 253–4
 - analysis 314
 - automated 412
 - comparative characteristics of resources 228, 229–31, 232–3
 - comparison with external data 162
 - completeness 247, 250–5
 - costs 341
 - de-identified 307
 - demographic 183, 192, 200
 - denominator 131
 - diagnosis 168–9
 - disease incidence 217–18, 234
 - dredging 123
 - drug claims to insurers 168
 - drug utilization 399
 - economic analysis 340–1
 - exposure 232
 - follow-up 318–19
 - genetic 201–2
 - identifiable 307
 - inpatient systems 436, 438
 - large simple trials 314
 - medical record databases 250–1
 - monitoring 318
 - outcome 146, 232–3
 - patterns 123
 - privacy 307
 - stratification 125
 - systematic analyses of automated 412
 - volume 401
- data collection 314, 315, 450
 - follow-up 318
 - medical devices 414
 - MedWatch program 102–3
 - meta-analyses 354
 - methods 243
 - non-spontaneous sources 82
 - pharmaceutical industry 80
 - prospective 22
 - retrospective 22
- data mining 109, 123
 - proportional reporting ratio 126–7
 - signal detection 109, 123, 128
 - vaccine safety signal detection 412
 - validation 125–6
- data protection 307
- data set
 - access to 307–8
 - de-identification 307, 308
 - limited 308
- data sources 256
 - accuracy evaluation 247–9
 - case-control studies 218–23
 - cohort studies 220
 - current 402
 - methodologic studies 247–8
 - secular trend analysis 215–18
- data validity 232–3, 239–57
 - computerized databases 250–5
 - database resources 250–5
 - disease 250
 - hospitalization 250
 - large simple trials 315
 - sensitivity/specificity 244
- database resources 74–5, 256
 - case-control studies 232
 - case-control surveillance 138, 148
 - cohort studies 232
 - comparative characteristics 228, 229–31, 233–4
 - computerized 71
 - confounding 270–1
 - data mining 109
 - data validity 250–5
 - exposure data validity 232
 - inpatient 429, 432–3
 - medication errors 428
 - new drugs 235–6
 - population-based 228, 232
 - relative cost 228
 - repository 66–7
 - representativeness 228
 - research questions 233–6
 - size 228
 - speed of studies 228
 - urgent questions 236
 - WHO 117–18, 123–6
- databases, automated 167–70, 173–209, 449
 - applications 170
 - data validity 250–5
 - design 392
 - General Practice Research Database 204–7, 209
 - Group Health Cooperative database 174–8, 207
 - HMO Research Network 208
 - Kaiser Permanente Medical Care Program 178–82, 208
 - Medicaid 188–92, 207
 - Netherlands Automated Pharmacy Record Linkage 196–8, 207
 - Saskatchewan health services database 192–5, 207
 - strengths 169
 - Tayside Medicines Monitoring Unit (MEMO) 198–203, 209
 - UnitedHealth Group 184–8, 208
 - weaknesses 169–70
 - see also* HMO Research Network
- Dear Health Care Professional letters 103, 107, 108
- debrisoquine 47
- decision support, computerized 405–6, 408
- deCode project 297
- defined daily dose 132, 402
- demographic change 64, 218

- demographic data
 - HMO Research Network 183
 - Saskatchewan health services database 193
 - Tayside Medicines Monitoring Unit 198
- denominator data 131
- depletion of susceptibles 36
- depression, diagnosis recall 241, 242
- descriptive studies 76, 474
- developing countries, pharmacovigilance programs 134–5
- development of drugs
 - cost 74
 - economic evaluation 334–6
 - epidemiology 77–8
 - process 334–6
 - regulatory/scientific advice 79
 - time span 74
- dexfenfluramine 40–1
- Diabetes Audit and Research in Tayside Scotland (DARTS) database 200
- diagnosis
 - administrative databases 253–6
 - coding systems 218
 - information validity 433
 - self-reported 250
 - terminology changes 218
- diagnostic data validity 168–9, 190
- diagnostic databases
 - General Practice Research Database (UK) 204
 - Medicaid 189, 190
 - outpatients 175, 181, 233
 - Saskatchewan health services database 194
 - Tayside Medicines Monitoring Unit 198, 207
- diaries, prescription 244
 - adherence to treatment 368
- diazepam 46
- dietary supplements 137
 - adverse event reporting 99
 - case–control surveillance 137, 138
 - information need 169–70
- diethylene glycol toxicity 5, 93
- diethylstilbestrol 18, 19
- digoxin 38, 40, 49, 361
- diltiazem 49
- diphtheria–pertussis vaccine 261–2
- directly observed therapy 368
- disease/disease state
 - alteration by genetic polymorphisms 290
 - background incidence 132
 - benefit–risk impact 82
 - causal pathways 290
 - claims databases 168
 - data validity 250
 - definition 262
 - dread 60
 - incidence 74
 - data 217–18, 234
 - low outcome 234
 - index of suspicion 218
 - lifestyle impact 250
 - nondifferential misclassification 246, 472
 - orphan 79
 - pharmacodynamic effects 39–40
 - phenotypic expression 294
 - postmarketing surveillance 59
 - prevalence 74
 - reportable 218
 - typically drug-induced 130
- distribution of drug
 - control 86
 - systemic 43–4
- distribution phase 41–2
- DNA 288
 - collection 142
 - sequence variation 295
- doctors *see* physicians
- dosage of drug
 - effect modification 267
 - natural experiments in dose ranging 376
 - postmarketing reductions 372
- dose ranging, natural experiments 371, 376
- dose–response relationship 16–17
- dosing errors 429
 - frequency 428
 - patient 368, 371
 - pharmacodynamic correlates 371
- dosing histories 370, 376–7
 - electronic monitoring 450
- dosing regimen
 - deviations from 376
 - forgiveness 373–4
 - omissions 369, 375–6
 - optimal 371–2, 373
 - perfect use 373
 - selection methods 372–3
 - typical use 373
 - variable 376–7
- doxycycline 372
- dread diseases 60
- drug(s) 36–7, 473
 - active component 36
 - alternatives 59, 60
 - channeling 394
 - co-ingestion with chelating agents 43
 - comedication 268
 - comparisons 15
 - concentration 49
 - continuity of action 374–5
 - counterfeit 135
 - electronic medication event monitoring 370
 - gene effects on responses 288–9
 - inactive ingredients 36
 - mechanisms of action 37
 - misuse 61
 - new 235–6
 - non-prescription 52, 145
 - outcome of interest 36
 - patient knowledge 240
 - plasma concentration 369
 - plasma half-life 370
 - post-dose duration of action 374
 - potency 37
 - effect modification 267
 - prescription 52
 - quality regulation 79
 - reformulation 36–7
 - statistics 400, 401
 - systemic distribution variation 43–4
 - targeting of therapy 450
 - undertaken 367–8
 - user fees 69
- drug–ADR combinations 124–5, 128
- drug approval process 9
 - policy 448
- drug database design 394
- drug dictionary, case–control surveillance 142
- drug–disease associations, case–control surveillance 141
- drug–drug interactions 40–1, 125, 126, 436
 - drug metabolizing enzymes 48–9
 - exposures 429
 - polypharmacy 432
 - protein binding 43
 - renal elimination 45
- Drug Efficacy Study Implementation (DESI) process 6
- Drug Epidemiology Unit 6
- drug-event pairs 71
- drug holidays 368, 378–9
- drug–laboratory errors 429–30
- drug metabolizing enzymes 48–9, 422

- drug response
 - adaptive 38–9
 - genetic determinants 38
 - pharmacokinetics 41–50
 - variability 38–41
- Drug Safety Research Trust (UK) 6, 155
- Drug Safety Research Unit (UK) 152
- drug transporters 43–4, 49
- Drug Use Evaluation (DUE) Program 436
- drug use evaluation (DUE) studies 400
- drug utilization 399–400, 473
 - see also* use of medication
- drug utilization review 399, 408–9, 439
 - criteria 408
 - validity 409–10
- effectiveness evaluation 406
- exceptions 408
- methodologic problems 408
- prospective 408, 478
- research 409
- retrospective 408, 478
- drug utilization review programs 400, 408–9, 449
 - hospital 408–9
 - lack of effect 408
 - outpatient
 - prospective 408–9
 - retrospective 408, 409
- drug utilization studies 6, 10, 400–3, 448, 475
 - classification systems 403
 - clinical problems 400–1
 - data 216–17, 400
 - indicator-based approach 402
 - measurement units 402–3
 - methodologic problems 401
 - patient outcome measures 403
 - process measures of quality 403
 - qualitative 400, 478
 - quantitative 400, 478
 - research 400–1
- duration–response relationship 17
- ecological studies 19, 473
- economic analysis
 - blinding 341
 - cost data 341
 - data collection 340
 - intermediate endpoints 341
 - measurement 341
 - methodologic problems 336–8
 - modeling 341
 - multicenter evaluations 340
 - patient participation 340
 - perspective 338
 - Phase III study data 334
 - sample size 340
 - statistical tests 341
 - study design 339–41
 - types 336–7
 - uncertainty 342
 - usual care arm 339
- economic evaluation 333–43
 - dimensions 336–8
 - drug development 334–6
 - generalizability 342
 - research 333–4
- economics of drug use 10–11, 52
 - see also* pharmacoeconomics
- educational materials, printed 403
- efavirenz 106–7
- effect measures
 - applicability 327
 - beneficial 321–31, 473
 - bias 244
 - confounding 265
 - cost-effectiveness studies 329–30
 - minimum 26
 - nonexperimental studies 327–8, 329, 330
 - time-window 391
 - types 321
- effect modification 267, 275
 - age 266
 - confounding 266–7
 - dose 267
 - genetic variants 293
- effectiveness studies 321–30, 448, 474
- effects of drugs
 - antagonistic 40
 - beneficial 322–3
 - anticipated 322
 - unanticipated 322, 478
 - delayed 235
 - harmful
 - anticipated 321
 - unanticipated 321, 478
 - transient 383
- efficacy of drug 324, 475
 - clinical outcomes 82–3
 - data analysis 78
 - degree of 83
 - RCTs 330
 - regulation 79
 - vaccine studies 329
- efficiency studies 322
- elderly people
 - chronic use of drugs 134
 - HMO representation 177
 - pharmacodynamic effects of drugs 39
 - proportion of population 64
- electronic medication event monitoring (eMEM) 370
- electronic monitoring 450
- elimination of drug, variation 44–5
- elimination phase, terminal 42
- elimination rate constant 42
- endpoints, intermediate 341
- enrolled person-time 191–2
- enzymes 37
 - drug metabolizing 48–9, 422
 - induction 48, 49
 - inhibition 49
- EpicCare® 176, 180–1
- EPICURE program 390
- epidemiology 4–5, 35–6, 473
 - AER report evaluation 105
 - descriptive studies 71
 - development of drugs 77–8
 - ethics 301–9
 - guidelines 302
 - pharmaceutical industry 69–74
 - post-licensing 80
 - practice 302
 - product planning/development 73–4
 - public understanding 75
 - quantitative methods 246–7
 - randomized 202
 - regulations 302
 - study designs 17–21
- epipodophyllotoxins 72
- EPITOME program 390
- equity in regulatory action 85
- errors
 - administration route 428
 - allergy 428
 - reduction 427
 - system-based interventions 429
 - type I 14–15, 26, 294, 296–7, 479
 - type II 14–15, 26, 294, 297, 479
 - see also* dosing errors; measurement error; medication errors
- erythromycin 49
- estrogens 327
- ethics
 - adherence to treatment 377
 - epidemiology 301–9
 - genetic testing 298
 - prescribing behavior RCTs 403
 - research 77, 301–9
 - Tayside Medicines Monitoring Unit (MEMO) 201
- ethics review boards 303
 - process 308

- ethinylestradiol 48
- ethnicity, confounding by 294–5, 297
- Eudranet 119
- European Medicines Evaluation Agency (EMA) 68, 119
 - international reporting standards 128–9
- European Union
 - Directive on data protection 307
 - pharmacovigilance system 119
- Euroqol (EQ5) 350
- evidence hierarchy 76
- Excellence in Pharmacovigilance model 76
- expert medical reviewers 70
- exposure 36
 - causal relationship 278
 - changes 384
 - defining criteria 270
 - dichotomous 293–4
 - differential misclassification 263
 - drugs in hospital 432–3
 - genetic polymorphism interactions 294
 - hospital stays 432
 - improper definitions 394
 - measures 244, 274, 449
 - misclassification 244
 - nondifferential 243, 245–6, 264
 - status 412
 - multiple variables 32
 - prediction 325–6
 - report accuracy 248
 - variations 376
- exposure–genotype interaction 143
- extension phase 78
- face validity 346, 475
- factor V Leiden mutation 290
- fail-safe *N* 357
- Federal-Wide Assurance (FWA) 303
- feeling thermometer 350
- fenfluramine 40–1
- fenoprofen 360
- fenoterol 6, 269
- Finland, medical record linkage system 224
- first line therapy 83
- first pass metabolism 48
- fluoxetine 6
- Food, Drug and Cosmetic Act (1938) 5, 93
- Food and Drug Administration (FDA) 4, 5
 - NDA fees 7
 - regulatory action 107
 - spontaneous reporting of adverse reactions 91–115
- foods, drug metabolism effects 49
- forgiveness, dosing regimen 371, 373–4
- formulation of drugs 58–9
- Friuli Venezia Giulia (FVG) database (Italy) 225
- funding
 - clinical pharmacology 66
 - training 447–8
- funnel plot 357
- gemfibrozil 133
- gene(s)
 - discovery 292, 295
 - drug response effects 289–91
 - mutations 38
- gene–drug interactions 290–1
 - case–control studies 295
 - causality 290
 - multiplicative 293
 - pharmacodynamic 289–90
 - pharmacokinetic 289
- General Practice Research Database (GPRD) 67, 204–7, 209, 252
 - comparative characteristics 228, 232, 233
 - completeness 205–6
 - computer complexity 206
 - data collection 204–5
 - diagnoses 204
 - medical attention 235
 - medical records 206
 - access 206
 - population-based data 205
 - prescriptions 204
 - quality assessment 255
 - size 205
 - validity 204–5
- general practitioners (GP) 152
 - Netherlands 196–8
 - Scotland 198, 199–200
- generalizability
 - economic evaluation 342
 - medication errors 428
- generalized linear model 342
- genetic data, phenotypic linkage 201
- genetic factors
 - clinical measurement 448–9
 - metabolism of drugs 47–8
- genetic polymorphisms 38, 163
 - β_2 -adrenergic receptors 289–90
 - case–control surveillance 146
 - disease state alteration 290
 - exposure interactions 294
 - metabolism of drugs 46–7, 422
 - prevalence 294
- genetic testing 298
- genetic variability 288–9
 - phenotypic expression of disease 294
- genetic variants 293–4
- genome-wide scans 292, 295
- genomic controls 297
- genotype 163, 448
 - drug exposure 143
 - drug metabolism alteration 290–1
 - orphan 298
- glibenclamide 133
- global introspection 278, 281
- Global Medical Device Nomenclature (GMDN) 415
- glomerular filtration 44
- glomerular filtration rate (GFR) 45
- glucocorticoids 203
- glutathione 48
- good clinical practice (GCP) 77
- good epidemiological practice (GEP) 201, 447
- G-protein-coupled receptors 37
- green form questionnaires 152, 157, 158
 - meloxicam 158, 159
- prompting effect 156
- Group Health Cooperative clinical trials 223
- Group Health Cooperative databases 173–8, 207, 251
 - accessibility 176
 - cancer surveillance 175
 - cause of death data 175
 - claims databases 176
 - clinical information systems 176
 - community health services 175
 - comparative characteristics 228, 233
 - completeness 176–7
 - enrollment 174
 - hospitalization 175
 - immunizations 175–6
 - inpatient 254
 - laboratory data 175
 - outpatient visits 175
 - pharmacy files 175
 - quality of data 176
 - radiology data 175
 - rare events 176
 - strengths 176

- Utilization Management/Cost Management Information System 175
- weaknesses 176–8
- Guillain-Barré syndrome 131
- haloperidol 191–2
- haplotypes 295
- harms of treatment 50–1
- Harvard Pilgrim Health Care 252
 - database 254
- hazards
 - increase 82
 - risk levels 61
- health care
 - access with pharmacogenetic testing 298
 - delivery models 67
 - practice 64–5
 - quality measurement in delivery 67
 - transactions 65
- Health Evaluation through Logical Processing (HELP) System 225
- Health Insurance Portability and Accountability Act (HIPAA) 190
 - confidentiality 256
 - medical data privacy 307
 - Privacy Rule 255, 256, 307
- health maintenance organizations (HMOs) 174–5, 334
 - elderly people 177
 - network models 179
 - out-of-plan care 177
 - poor people 177
 - uninsured people 174, 184
- health profiles 350, 475
- health-related quality-of-life (HRQL) 345–6, 350, 475
 - anchor-based approach 348
 - clinical problems 346
 - distribution-based approach 348–9
 - instruments 346–51
 - methodologic problems 346
 - research 346
- health surveys 217–18
- Health Utilities Index 350
- Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS) 417
- heart disease 407–8
- heart failure 335–6
- Helsinki Declaration (1964) 303
- herbal supplements 137, 449
 - information need 256
- HIV infection 82–3
- HMO Research Network 66, 178, 182–4, 208
 - Black Box Warnings 184
 - Center for Education and Research on Therapeutics 182
 - comparative characteristics 228, 233
 - external registry linkages 180–1
 - medical records 183
 - member health plans 182–3
 - membership status 183
 - new drugs 235
- homeostatic regulation 39
- hormone replacement therapy (HRT)
 - case-control surveillance 138, 148
 - combination 8, 21
 - selection bias 220
- hospital(s)
 - admissions 425
 - adverse drug reaction monitoring 436–8
 - care outcomes 432
 - diagnosis information validity 433
 - drug exposure 432
 - drug information validity 433
 - drug use 432–3
 - evaluation programs
 - intensive surveillance 434–5
 - staff participation in studies 434
 - see also* inpatients; outpatients
- Hospital Discharge Survey 218
- hospital pharmacoepidemiology 432–9, 440
 - clinical problems 428–30
 - methodological problems 430–1
- hospitalization 203
 - administrative databases 253–6
 - data validity 250
 - self-reported 250
- human growth hormone (hGH) 72
- human subjects, research 301–3
- hydralazine 133
- hydroxylation polymorphism 47
- hypertension, unresponsive 378
- hypotension, postural 39
- hypothesis generation 58–9, 76, 162, 233, 476
 - effect modification 267
 - postmarketing drug surveillance 110–11
 - signal evaluation 122–3
- hypothesis-strengthening studies 233–4, 476
- hypothesis testing 58, 167
 - case-control surveillance 138, 142–3
 - large simple trials 316
- randomized controlled trials 312, 314
- studies 234, 476
- ibuprofen 311–12, 360
- Icelandic people, genetic data 297
- immortal time bias 383, 394
- impact analysis concept 81
- implants, long-term 414
- IMS Disease Analyzer 224–5
- IMS HEALTH 216, 224–5
- incidence
 - bias 476
 - density 160
 - predetermined 31
 - rate 32, 476
 - rate ratio 160
- incomplete penetrance 422
- indication
 - confounding 265–6, 271–2, 312–13, 324–5, 326, 474
 - birth defects 418
 - measure 326
- indomethacin 6
- induction therapy in renal transplantation 360–1
- information
 - coherence with 15–16
 - framing 50–1
 - sources 135, 243
 - see also* bias, information
- information technology 429
- informed consent
 - case-control surveillance 138
 - ethical requirement 309
 - genetic data of Icelandic people 297
 - large simple trials 317
 - pharmacogenetic testing 298
 - research 304–5, 306–7
 - waiving 304–5
- inpatients
 - data systems 433, 436
 - databases 429, 432–3
 - drug use 233, 235, 430
 - health status 430
 - information link to outpatient information 433, 438
 - intensive drug monitoring 434–5
 - records 431
 - referral bias 434
 - see also* hospital(s)
- Institute of Medicine (IOM), vaccine safety review 411
- institutional review boards (IRBs) 301, 303
 - patient authorization waiver 308

- insulin, human 6
- intention-to-treat analysis 318
- interaction 265
 - effect modification 268–9
- International Classification of Diseases (ICD) coding 253
- International Classification of Diseases–Ninth Revision–Clinical Modification (ICD-9-CM) 168, 183, 190
- international collaboration 64
- International Conference on Harmonisation (ICH)
 - guidelines 79
 - postmarketing surveillance 100–1
- International Society for Pharmacoeconomics 64, 447
- interpretability, HRQL instruments 347
- interrupted time-series 405, 476
- intervention, purpose 82
- intussusception 412
- Investigational New Drug Application (IND) 5–6
 - causality assessment 280–1
- Iowa Drug Information System (IDIS) 403
- isoniazid 46–7
- isotretinoin 6, 61, 418
- Joint Commission on Accreditation of Healthcare Organizations 436
- Joint Commission on Prescription Drug Use 6
- Kaiser Permanente Medical Care Program 178–82, 208, 254
 - administrative database 179–80, 254
 - cancer incidence data 180
 - clinical database 179–80, 254
 - comparative characteristics 228, 232, 233
 - death certificates 180
 - electronic medical records 180–1
 - hospital discharge records 180
 - laboratory test records 180
 - member dropout rates 181
 - membership 179, 181
 - multiphasic health check-up 180
 - outpatient diagnostic databases 181
 - outpatient visits 180
 - pharmacy databases 179–80
 - research centers 178–9
 - socioeconomic status data 180
- Kefauver–Harris Amendments (1962) 5–6, 93, 323
- ketorolac 58, 438–9
 - pharmacy-based study 221, 222
- knowledge base 64
 - deficits 65
 - improvement 66
- labeling 107, 113–14
- laboratory results data 431
- large simple trial (LST) design 71–2, 314–18
 - analysis 318
 - compliance 317
 - confounding 316
 - cooperative population 317
 - data 315
 - monitoring 318
 - feasibility 316–17
 - follow-up 317–18
 - informed consent 317
 - logistics of conducting 317–18
 - objectivity 316–17
 - outcome measurement 316–17
 - power of study 315, 316
 - registration documents 317
 - research question 315
 - sample size 315
 - simple treatments 316
 - subgroup analysis 318
 - uncertainty principle 315–16
 - validity 317
- latent interval analyses 143
- legal issues
 - genetic testing 298
 - postmarketing surveillance 58
 - prescribing behavior RCTs 403
 - teratogens 419–20
- liberalism 304
- licensing 79–80
 - safety of drug 78–9
- life cycle of drugs 77–86
- life expectancy, disability-free 64
- LifeLink™ 216
- lifestyle impact on disease 250
- linezolid 113–14
- linkage disequilibrium 295
- literature search for meta-analyses 355
- lithium 49
- liver disease
 - drug sensitivity 40
 - metabolism of drugs 47–8
- liver failure, acute 187–8, 426
- logistic regression, conditional 389
- long latency outcomes 72
- longitudinal histories 186
- lower limit factor 31
- lung cancer, cigarette smoking 15, 16, 17, 19
- lung disease, drug sensitivity 40
- Lyme vaccine 7
- managed care organizations (MCOs) 373
- Manitoba data files 219
- Manitoba Health Services Insurance Plan 219
- Mantel–Haenszel odds ratio 272, 356, 359, 392
- markers
 - interactions with medication 293–5
 - intermediate 450
- market repositioning of drug 56
- marketing 57–9
 - postmarketing drug surveillance studies 223
 - timing 57
- matching, study design 270
- Mayo Clinic system 219
- measurement error
 - correcting measures of association 246–7
 - differential 242
 - nondifferential 242
 - quantitative indices 243–5
 - research 242–3
- measurement process standardization 270
- MedDRA 129
- media
 - informing 67
 - misinterpretation of studies 75
 - safety issues 85
- Medicaid
 - claim validity 253–4
 - comparative characteristics 228, 233
 - data
 - sources 188
 - validity 190
 - databases 188–92, 208, 252, 254–5
 - delayed drug effects 235
 - diagnosis accuracy 255
 - diagnostic terminology 190
 - diagnostic validity 190
 - enrolled person-time 191–2
 - enrollment 177, 191
 - generalizability of population 190
 - lack of benefits 65
 - medical attention 235
 - new drugs 235
 - overrepresented populations 189

- program 188
- recipients 188
- UnitedHealth Group
 - members 185
- medical care
 - costs 333–4
 - reimbursement 342
- medical devices
 - care claims 415
 - clinical problems 414
 - data collection 415
 - definition 414
 - exposure assessment 415
 - medical records 415, 417
 - methodologic problems 414
 - national population exposure
 - assessment 415
 - public health impact 415
 - registries 415
 - safety surveillance 416
 - studies 414–24, 438
 - hypotheses 412
 - techniques 417–8
 - surveys 415, 416
- Medical Outcomes Study instruments 350
- medical record databases 169, 180–1
 - data 251–2
 - use 256
- medical records
 - abstractions 186
 - completeness 247
 - confidentiality 256
 - electronic 180–1, 183, 429
 - General Practice Research Database 205, 206
 - inpatients 431
 - linkage
 - in Netherlands 196–7
 - system 224
 - medical devices 414, 415
 - organization 434
 - permission for use 304
 - research 306–7
 - retrieval 247
 - Saskatchewan health services database 194
 - Tayside Medicines Monitoring Unit 198
 - validation 186, 247
- Medicare
 - data structure 189
 - diagnoses 189
 - drug benefits 232, 448, 450
 - UnitedHealth Group members 185
- medication errors 92, 427–31, 476
 - chart review 428
 - claims-based evaluations 428
 - clinical problems 428–9
 - data sources 430
 - database resources 428
 - generalizability 430
 - handoffs 428
 - information bias 430
 - methodological problems 430–1
 - patient safety 425–6
 - safety theory 427
 - sample size 430
 - types 426
- medication event monitoring, electronic (eMEM) 370
- Medicines and Healthcare Products
 - Regulatory Agency (MHRA) 76, 129
- Medicines Monitoring Unit (MEMO) *see* Tayside Medicines Monitoring Unit (MEMO)
- medroxyprogesterone, depot 6
- MedWatch Adverse Event Reporting
 - Form 94, 95–8
- MedWatch program 102–3, 108
- MedWatch to Manufacturer Program (MMP) 103
- meloxicam 158, 159, 160
- membership databases
 - HMO Research Network 183
 - Kaiser Permanente Medical Care Program 178, 182
- memory
 - autobiographical 240–2
 - organization 241
 - retrieval 241
- MEMS[®] Monitors 371
- meta-analyses 76, 353–63, 476
 - adverse effects studies 357–60
 - applications 357–62
 - Bayesian approach 356
 - clinical problems 354
 - combinability of studies 354, 356–7
 - conclusions 356
 - cost implications 361–2
 - cumulative 361–2, 363
 - data
 - abstraction bias 354–5
 - collection 355
 - definition 353–4
 - diversity of studies 356–7
 - group-level 362
 - heterogeneity 356–7, 362–3
- inclusion/exclusion criteria 355
- individual-level data availability 362
- induction therapy in renal transplantation 360–1
- literature search 355
- methodologic problems 354–5
- new indications for existing therapies 360–1
- nonexperimental studies 356–7
- prospective 363
- publication bias 354
- purpose 355
- quality of original studies 354
- recommendations 356
- research 354–5
- selection of therapies 361
- statistical analyses 356
- steps 355–6
- time saving 361–2
- meta-regression techniques 362
- metabolic reactions 45–6
- metabolism of drugs
 - alteration by genotype 290–1
 - disease effects 47–8
 - genetic factors 46–7
 - slow 448
 - variation 45–9
- metabolites, active 48
- methadone 18
- methodologic advances 448
- metamide 18
- metoprolol 379
- minimal important difference (MID) 348, 349
- minimal risk 304
- minimum effect 26
- misclassification
 - bias 265, 476
 - differential 265
 - exposure 244
 - nondifferential 243, 245–6, 264
 - status 412
 - nondifferential 147, 243, 245–6, 264, 472
 - Stevens–Johnson syndrome 248
 - variables 243
- mitoxantrone 110–11
- modeling 271, 325
 - continuous-benefit effect
 - model 341
 - economic analysis 341
 - generalized linear model 342
 - one-time effect model 341
 - pharmacokinetic 370–1

- molecular pharmacoepidemiology
 - 287–98, 476
 - cost-effectiveness 298
 - methodological problems 292–5
 - research 291–2
 - type I errors 294, 296–7
 - type II errors 294, 297
- monitoring centers, national 119–20
- morbidity data 217
- mortality data 217, 218
- Multi-item Gamma Poisson Shrinker (MGPS) method 70
- multicenter studies, economic evaluation 340
- multiple regression 325
- multiple sclerosis 392–3
- Multivariate analysis 272
- predictive power 342
- myelo-optic-neuropathy, subacute 6
- myelosuppression 113–14
- myocardial infarction 324

- naloxone 18, 325
- named patient use 78
- naproxen 360
- National Ambulatory Medical Care Survey 74
- National Birth Defects Prevention Study 421
- National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program 175
- National Center for Health Statistics (NCHS) 216, 217
- National Childhood Encephalopathy Study (NCES) 261–2
- National Death Index (US) 217
- National Disease and Therapeutic Index™ 216, 217
- National Institute of General Medical Sciences (NIGMS) 452
- National Institutes of Health (NIH) 451, 453
 - biomedical research 75
 - Roadmap 67
 - training funding 448
- National Prescription Audit Plus™ 216
- National Sales Perspective™ 216
- Netherlands
 - automated pharmacy record linkage 196–8, 209, 228, 252
 - validation studies 255
 - community pharmacy system 196
 - general practice 196
 - medical record linkage 196–7
- neural networks, Bayesian approach 123–5
- neural tube defects 418
- New Drug Applications (NDA) 92
 - fee 7
- nifedipine 43
- non-prescription medications 52, 145
- non-steroidal anti-inflammatory drugs (NSAIDs)
 - adverse events 311–12, 358–60
 - angina 284
 - case-control surveillance 143–4
 - colorectal cancer risk 263–4
 - database resources 236
 - gastric ulceration 38–9
 - recall accuracy 248–9
 - renal function 40, 45
 - study types 236–7
- nondifferential misclassification 246, 476
- nonexperimental studies 327–8, 329, 330, 448, 476
 - meta-analyses 356–7
 - uncertainty in risk assessment 384
- nonparametric bootstrap methods 342
- number needed to harm 83
- number needed to treat 83, 349
- numerator data 132–3

- obesity, young people 64
- objectivity in regulatory action 85
- observational studies 71–2, 74, 476
 - causal relationships 312
 - confounding by indication 313
- oculomucocutaneous syndrome 6
- odds ratio 20, 143, 147, 476
 - case-time-control 391
 - combination 356
 - confounding 389
 - reporting 127
 - stratification 272
- Office of Drug Safety (ODS) 92, 93
- Olmsted County medical records 219
- omeprazole 49
- one-step method for meta-analysis
 - statistics 356
- one-time effect model 341
- opiates 40
- opinion leaders 406
- oral contraceptives 18, 19, 21
 - case-control study 242–3
 - case-control surveillance 144, 145
 - combined 156
 - confounding 270
 - dosing regimen 371
- ethinylestradiol 48
- meta-analysis of nonexperimental studies 356–7
- perfect use 373
- post-dose duration of action 374
- thromboembolic disease 133
- typical use 373
- unsuspected associations 144
- orphan medicines 79, 298
- osteoporotic fractures 327–8
- outcome
 - data 146
 - validity 232–3
 - defining criteria 270
 - differential misclassification 264
 - expected incidence 26
 - hospital care 432
 - of interest 36
 - measurement 85
 - large simple trials 316–17
 - measures 243
 - multiple variables 32
 - nondifferential misclassification 264
 - patient of physician prescribing 404
 - patient-reported 201
 - pregnancy 162
- outcomes research 322
- outpatients
 - diagnostic databases 175, 181, 233
 - information 430, 435
 - prescription encounters 189
 - therapeutic transaction 65
 - see also* hospital(s)
- over-the-counter (OTC) drugs
 - adverse events 94
 - information need 169–70, 177, 202, 256
 - teratogenic 418
- package inserts 129
- pancreatitis 114
- passive diffusion 42
- patient(s)
 - authorization for data use 308
 - behavior 52
 - comparison groups 339
 - demand 50
 - dosing errors 368, 369
 - expectations 50
 - handoffs 428
 - hospitalized 429–30
 - information sources 135
 - knowledge of drugs 240
 - preferences 61
 - responsibility for medication 368

- rights 305
- selective consent 306–7
- self-monitoring surveillance studies 221
- therapeutic regimen recall 239–40
- welfare 305
- patient identification
 - health services number 192, 194
 - Tayside Medicines Monitoring Unit 198, 201
- patient safety 428
 - movement 8
 - societal focus 66
- peer review 83
- penicillin 18
- periodic safety update reports (PSUR) 80
- person-time at risk 186
- personnel resources 451–2
- P-glycoprotein 43–4, 49
- pH of drugs 44
- pharmaceutical industry 51–2, 68–76
 - causality assessment uses 280–1
 - data collection 80
 - epidemiology 69–74
 - globalization 80
 - mandatory reporting of adverse events 94, 99
 - orphan drug investment 298
 - pharmacoepidemiology units 448, 452
 - regulatory agency relationship 68–9
 - science relationship 75
- pharmacodynamics 37–8, 477
 - age effects 39
 - disease state effects 39–40
 - drug response variability 38–41
 - drug–drug interactions 40–1
 - gene–drug interactions 289–90
- pharmacoeconomics 10–11, 51, 333–43, 449
 - prospective study design 339–42
 - research trends 334
 - third-party payers 342
- pharmacoepidemiology
 - contributions 9–11
 - definition 3–5, 477
 - history 5–9
 - resources 74–5
 - study interpretation 75
- pharmacogenetics 163, 288–9, 448–9, 477
 - testing 298
- pharmacogenomics 38, 288–9, 477
- pharmacokinetic compartments 41, 477
- pharmacokinetic model 370–1
- pharmacokinetics 4, 37–8, 477
 - drug response 41–50
 - gene–drug interactions 289
 - mathematical parameters 41–2
 - variability 49–50
- pharmacology 4, 477
- pharmacotherapeutics 35, 477
- pharmacovigilance 477
 - aims 119
 - plan 79–80
 - post-licensing 80
 - programs in developing countries 134–5
 - routine 80
 - scope 118–19
 - unidentifiable data 307
- Pharmacovigilance Planning guideline 79
- pharmacy-based surveillance studies 220–2, 235
- pharmacy databases
 - General Practice Research Database 204
 - Group Health Cooperative databases 174–5
 - Kaiser Permanente Medical Care Program 178–9
 - Medicare 189
 - Netherlands automated pharmacy record linkage 196–8
 - patient files 197
 - Saskatchewan health services 192
 - Tayside Medicines Monitoring Unit 198–9
 - UnitedHealth Group 185, 187
- pharmionics 367–8, 477
 - terminology 368–9
- PHARMO system 196
- Phase I testing 9
 - metabolic reactions 45, 46
- Phase II testing 9
 - metabolic reactions 45–6
- Phase III testing 9
 - economic data 334
- Phase IV studies 18, 31, 56
 - risk management 423
- phenolphthalein 145
- phenylbutazone 6, 46
- phenylpropanolamine 7, 108–9
- phenytoin 49, 419
- physicians
 - ad hoc* cohort studies 222
 - behavior studies 404
 - clinical feedback 406
 - diagnosis 239
 - identification 186
 - pressures 135
 - services data 193
 - see also* prescribing, physician
- pioglitazone 164
- piroxicam 58
- placebo controls 10, 314
- plasma concentration of drug 42
- plasma concentration time curves 41
- point estimate of association, measurement error 245–6
- poison control centers 450
- Poisson-distributed variables, confidence intervals 471
- policies, counterproductive 65–6
- policy decisions 450
- policy makers, informing 67
- polypharmacy 432, 433
- poor people, HMO representation 177
- population
 - admixture 294–5, 297
 - being treated 82
 - elderly 64
 - exposed 4
 - exposure estimation 112
- post-approval studies 74
- post-licensing studies 80
 - design 77
- postmarketing drug surveillance 7, 8, 9–11, 18, 477
 - beneficial effects 323
 - clinician contributions 110
 - cohort studies 20, 137–8
 - cost-effectiveness 110
 - data collection 102–3
 - databases 167–8
 - disease state 59
 - hypothesis generation 110
 - ICH guidance 100–1
 - incidence rate 32
 - legal issues 58
 - marketing emphasis 223
 - monitoring 152
 - patient self-monitoring 221
 - pharmacy-based studies 220–2
 - prescriber behavior 57
 - randomized controlled trials 223–4
 - rare adverse reactions 33
 - recommendations 152
 - reporting 151–2
 - risk management 421
 - scale 110
 - signal detection 110
 - strengths 109–11
 - toxicity information 57–8
 - weaknesses 111–12

- postmarketing, risk management 423
- potassium chloride formulation trials 223
- practice
 - epidemiology 302
 - health care delivery 64–5
- practolol 6
- pravastatin 195
- prazosin 10, 18, 31, 56
 - postmarketing surveillance 223
- pre-approval period, risk management 69
- pre-licensing stage 77–9
- PRECEDE model 406
- pregnancy
 - antidepressant safety 177–8
 - drug evaluation 73
 - drug use 323
 - exposure registry 73
 - medication errors 428
 - outcomes 162
 - see also* birth defects, drug-induced; teratogens
- premarketing studies 9, 10, 71, 312
- efficacy 322
- prescribed daily dose 132, 403
- prescriber behavior 50–2
 - adherence to treatment guidelines 404
 - cost containment 51
 - modification 401, 403
 - patterns 51
 - pecuniary interests 51
 - pharmaceutical industry influence 52
 - physician prescribing 10
 - postmarketing drug surveillance 57
- prescribing
 - behavior modification 403
 - budgets 51
 - conceptual framework 406
 - coverage limitations in Medicaid 190
 - databases 189, 192, 198–9
 - errors 404, 477
 - General Practice Research Database 204
 - internal validity 405
 - intervention effectiveness 405
 - physician 10, 436–7
 - clinical feedback 406
 - clinical problems 404
 - evaluation 404
 - improving 404–7
 - methodologic problems 405–6
 - research 402–4
 - prescriber data 199
 - reasons for 267
 - records 175
 - trend monitoring 74
 - unit of analysis 405
- prescription(s) 152
 - indication for 267–8, 273–4
 - information for medication errors 428
 - number 400
 - sequence analysis 394
 - unclaimed 251
 - unique identifier number 199
- prescription diaries 244, 369
- Prescription Drug User Fee Act (US, 1992) 7, 69
- Prescription-Event Monitoring (PEM) 6, 151–64
 - applications 157–8, 159, 160–3
 - background effects 163
 - data analysis 153
 - data comparison with external data 162
 - database resources 234, 235
 - event analysis 158, 160
 - event data collection 153
 - event rates 160
 - GP response rate 153, 155
 - incidence density 160
 - long latency adverse reactions 160–2
 - modified studies 163
 - pregnancy outcomes 162
 - process 152–3, 154, 163
 - prompting effect 156
 - reasons for stopping drug use 157, 160
 - risk management 163–4
 - signal detection 157, 158, 159, 160–2
 - signal generation 160
 - strengths 155–7
 - weaknesses 157
- Prescription Pricing Bureau 223
- press, informing 67
- presystemic clearance 48
- prevalence 476
 - bias 264, 476
 - low exposure 234–5
- PREVEND (Prevention of Renal and Vascular End-stage Disease) study 196–7
- privacy 303, 477
 - authorization waiver 308
 - data 307
 - inpatient data systems 436
- Privacy Board 308
- probenecid 45
- products
 - name recognition 57–8
 - planning 73–4
 - quality problems 92
 - withdrawals 107
- Programme for International Drug Monitoring (WHO) 117–36
- propensity scores 271, 273, 325–6, 448, 477
- proportional reporting ratio (PPR) 70, 126–7, 412
- propranolol 38, 379
- prospective studies 22, 478
- protease inhibitors 370–1
- protein binding 43
- protocol-induced testing 338–9
- protopathic bias 262, 264, 478
- public
 - informing 67
 - understanding of epidemiology 75
- public health 76
 - protection 79
- publication bias 474
 - meta-analyses 354, 357
- public–private partnerships 68
- Pure Food and Drug Act (1906) 5, 93
- p*-values 20, 21, 477
- quality-of-life measurements 345–51, 475
 - anchor-based approach 348
 - distribution-based approach 348–9
 - instruments 346–51, 474
 - discriminative 347
 - disease-specific 350–1
 - evaluative 347
 - generic 349–50
 - potential use 349–51
 - score interpretation 347
 - specific 349, 350–1, 479
 - taxonomy 349–51
- Quality of Well-Being Scale 350
- quasi-experimental designs 405
- questionnaires 241
- random-effects models 356
- randomization
 - by cluster 409
 - confounding control 268, 313
 - large simple trials 314–15
- randomized controlled trials (RCT) 9, 19, 21, 22, 71, 76, 311–19
 - blinding 313–14
 - classic 313–14
 - confounding control 268–9
 - control groups 314
 - data analysis/collection 314

- definition 478
- double-blind 223–4
- efficacy assessment 334
- hypothesis testing 313, 314
- limitations 314
- new drugs 235
- orphan medicines 79
- postmarketing drug surveillance 222
- premarketing 71
- prescribing behavior 403
- research 312
- results generalizability 314
- sample size 313
- randomized simplified trials 202
- rapacuronium 7
- recall
 - accuracy 243, 248, 249
 - NSAIDs 248–9
 - predictors 249
 - time lag 255
 - bias 146–7, 249, 263, 420, 478
- record linkages 185
- recruitment of patients 20
- referral bias 478
- inpatients 434
- Regenstrief Medical Record System 224
- regional centers 120
- registry data 218–19
- regression to the mean 325, 403, 474
- regulation
 - current requirements 94, 99–101
 - distribution of drug 86
 - efficacy of drug 79
 - epidemiology 302
 - history 5–9, 92–4
 - quality of drug 79
 - reporting requirements 94, 99–109
 - research 77
 - safety of drug 79
 - teratogens 419
 - use of medication 86
- regulatory action 85
- regulatory advice, development of
 - drug 79
- regulatory agencies 452–3
 - causality assessment 279–80
 - obligations 76
 - pharmaceutical industry relationship 69–70
 - principles 76–7
- regulatory approval, early 56–7
- regulatory authorities 79, 80
- regulatory studies 56–7
- reimbursement 168
- medical care 342
- reinforcements, positive 407
- reliability 478
 - HRQL instruments 346, 348
 - quantitative measurement 245
- renal dosing errors 429
- renal elimination
 - drug–drug interactions 45–6
 - variation 44
- renal function, NSAIDs 40
- renal transplantation, induction therapy 360–2
- renal tubular reabsorption, passive 44
- renal tubular secretion, active 44
- replication studies 296–7
- reporting odds ratio (ROR) 128
- reports/reporting
 - accuracy 248
 - bias 130
 - causality assessment 279–80
 - characteristics 163
 - computerized storage 93
 - 15-day 129
 - electronic 99–100, 101
 - evaluation 105–9, 121–2
 - feedback to reporters 122
 - handling 121–2
 - international standards 128–9
 - mandatory 91, 94, 97–8, 99, 120–1
 - modifications 101–2
 - publishers 281
 - quality 112
 - quantitative 128
 - requirements 120–1
 - for generation 129–30
 - schemes 119
 - selective 130
 - sources 121
 - underreporting 111, 132
 - voluntary 91, 95–6
 - see also* spontaneous reporting
- reproducibility 478
 - HRQL instruments 346–8
- research 474
 - adherence to treatment 365–8
 - beneficial effects 322–3
 - bioethics 302–5
 - clinical problems 239–40, 261–2
 - economic evaluation 334–6
 - ethics 77, 301–9
 - ethics board review
 - funding 451
- Health Insurance Portability and Accountability Act impact 256
- health-related quality-of-life 345
- human subjects 301–3
- informed consent 304–5, 307
- investment 64
- measurement error 242–3
- meta-analyses 353–5
- methodologic problems 240–7, 305–7, 313–14, 323–5
- minimal risk 304
- molecular pharmacoepidemiology 291–2
- outcomes 322
- participant confidentiality 75, 77, 303–4, 305–6, 307–8, 447
- randomized controlled trials 312
- regulation 77
- social services 309
- study designs 383
- validity 306
- research questions
 - beneficial effects of drugs 234
 - database resources 233–6
 - see also* hypothesis generation
- resources, pharmacoepidemiology 74–5
- responsiveness 478
 - HRQL instruments 346, 348
- retrospective studies 22, 478
- rhabdomyolysis 134
- rights of patients 305
- risk 55, 474
 - acceptability 59–61
 - action to reduce 85
 - adverse reaction types 93
 - balance with benefits 82–3
 - characterization 424
 - communication 186
 - comparative 71
 - evaluator perceptions 61
 - excess 21, 132
 - hazards 61
 - identification 424
 - impact 83
 - minimal 304–5
 - quantification 81–2
 - relative 20–1, 132, 474
 - versus* safety 59
 - summarizing 79
 - tolerance 59–61
 - voluntary assumption 61
- risk group identification 424

- risk management 8–9, 423–5, 440, 448
 - case reports 424
 - clinical problems 424–5
 - design 425
 - effectiveness 424, 426
 - engineering-based 425
 - evaluation 425
 - goal setting 425
 - methodological problems 424
 - Phase IV studies 424
 - postmarketing 424, 425
 - pre-approval period 69
 - Prescription-Event Monitoring 163–4
 - product usage 424
 - programs 425–6
- risk sets 386–8
- RNA 288
- rofecoxib 7
- root cause analysis 285
- rosiglitazone 164
- rotavirus vaccine 7, 412
- rule of threes 32, 69
- safety alerts 103
- safety data
 - clinical trials 78
 - collection by pharmaceutical industry 80
 - comparator-controlled 78–9
 - management 80
- Safety Information and Adverse Event Reporting Program (FDA) 102
- safety of drugs 58, 59
 - adverse event data collection 78
 - assessment 78–9
 - communication 85
 - degrees 59
 - evaluation 69
 - history of regulation 92–4
 - licensing 77–8
 - post-approval studies 74
 - postmarketing profile 70
 - regulation 79
 - specification 79–80
- safety theory 427
- salbutamol 349
- sample size 25–33
 - birth defects 418
 - calculation 26, 29, 32
 - case series 31–2
 - case-control studies 29, 30, 31, 32, 33, 464–70
 - clinical trials 69
 - cohort studies 26–7, 28, 29, 32, 455–62
 - economic analysis 340
 - large simple trials 314
 - medication errors 427
 - randomized controlled trials 313
 - vaccine studies 413
- sampling
 - strategies 383
 - two-stage 272, 391
- Saskatchewan health services database
 - 192–5, 209, 253–4
 - cancer services data 193
 - comparative characteristics 228, 233
 - data linking 194
 - demographic data 192
 - diagnostic information 195
 - eligible population 192
 - exposure data limitation 194
 - health services data 192–4
 - health services number 192, 194
 - hospital services data 193
 - medical records 194
 - physician services data 193
 - prescription drug data 192
 - reliability 194
 - validity 194, 255
- scientific advice, development of drug 79
- scientific inference 14, 478
- scientific method 13–14
- screening tests, cancer 329
- second line therapy 83
- secular changes, drug-induced birth defects 422
- secular trends analyses 19, 21–2, 473
 - data sources 215–18
- selection *see* bias, selection
- selective serotonin reuptake inhibitors (SSRIs) 83–5, 177–8
- sensibility 346, 478
- sensitivity 478
 - calculation 244
 - low 247
- sensitivity analysis 246, 256, 273, 479
 - cost-effectiveness 337
- Serious Adverse Drug Reaction (SADR) Reporting Proposed Rule 101–2
- serotonin 40–1
- SF-36 350
- Sickness Impact Profile (SIP) 350
- side effects, known 92
- signal(s)
 - automated 160
 - evaluation 122–8
 - false 130
 - generation 160
 - follow-up 237
- signal detection 70–1, 80–1, 122–8
 - adverse event reports 105, 107
 - data mining 109, 123, 127–8
 - evaluation 81–2
 - HRQL instruments 347
 - postmarketing drug surveillance 110
 - Prescription-Event Monitoring 157–8, 159, 160–2
 - prioritization 81
 - proportional reporting ratio 126–7
 - statins 134
 - Uppsala Monitoring Center system 124–5
 - vaccine adverse events 412–13
 - vaccines 411
- signal-to-noise ratio 347
- sildenafil 40, 152, 162
- silicone breast implants 6
- single nucleotide polymorphisms (SNPs) 288, 295
- Slone Epidemiology Center 6
 - case-control surveillance 228
- Slone Survey 216–17
- slow acetylators 46–7, 448
- small study effects 357
- smallpox vaccine 7
- social implications of genetic testing 298–9
- Social Security death tapes (US) 217
- social services research 308–9
- societies for pharmacoepidemiology 447
- sparteine 47
- specificity 479
 - of associations 16
 - calculation 244
 - low 247
- spirometry 347
- spontaneous reporting 70, 80, 81, 82, 91–115, 118, 478
 - case-control studies 219
 - limitations 129–30
 - mechanisms for adverse reactions 132–3
 - new drugs 235–6
 - risk estimation 132
 - risk group identification 132–3
 - strengths 129
- Spontaneous Reporting System (SRS) 93
- standard error of measurement (SEM) 349
- standard gamble 350
- standardized incidence rate (SIR) 389
- standardized mortality rate (SMR) 389

- statins 133–4, 195
- statistical analyses for meta-analysis 356
- statistical inference 14, 479
- statistical power, birth defects 421
- statistical significance 33, 476
 - economic analysis 341
- statistics 14
 - drug 398, 399
- steady state concentration 42, 479
- stem cell transplantation, autologous 334–5
- Stevens–Johnson syndrome
 - misclassification 248
- stratification, study design 272–3
- studies
 - approaches 227–37
 - interpretation 75
 - programs 441
 - reasons to perform 55–9
 - subjects 13–14
 - timing 55–62
- study designs 13–23, 383–95
 - clinical problems 384
 - confounding control 268–71
 - descriptive 21–2
 - epidemiologic 17–21
 - matching 269
 - methodologic problems 384–5
 - observational 21
 - pharmacoeconomics 339–42
 - post-licensing studies 77
 - regulatory agencies 76
 - research 382–3
 - restriction 269
 - sampling within a cohort 384–8
 - standardization 269, 270
 - stratification 272–3
 - two-stage sampling 272–3
 - within-subject 383, 391–4
 - see also named types*
- sulfanilamide elixir 5, 93
- Summary of Product Characteristics
 - package insert 129
- supraventricular arrhythmias 361
- suprofen 6, 133, 219
- suspected unexpected serious
 - adverse drug reaction (SUSAR) 78
- Sweden, data sources 217
- syndrome detection 125
- synergy index 296
- system-based interventions,
 - errors 429
- system changes, error
 - reduction 427
- tampons 60
- Tayside Medicines Monitoring Unit (MEMO) 198–203, 209, 252
 - admissions 203
 - birth cohort 200
 - clinical laboratory data 200
 - comparative characteristics 228, 233
 - confidentiality 201, 202
 - demographic data 200
 - diagnostic accuracy 202, 255
 - drug exposure data 202
 - episode of care 203
 - ethics 201
 - genetic data 201
 - good epidemiological practice 201
 - hospital data 199–200
 - inpatient admissions 255
 - medical records accessibility 202
 - patient access 202
 - patient identification 199, 201
 - patient-reported outcomes 201
 - population-based data 201
 - population size 202
 - prescribing data 198–9
 - primary care data 200
 - randomized trials 202
- temafloxacin 6, 113
- temazepam 46
- teratogenesis
 - class action 419
 - unknown risk 416
 - see also* birth defects, drug-induced
- teratogens 5, 6, 61, 418–23
 - alleged 419
 - high risk 418
 - legal issues 419–20
 - moderate risk 418
 - over-the-counter drugs 419
 - pregnancy outcomes 162
 - regulation 422
 - study of effects 73
- terfenadine 6
- tests
 - one-tailed 26, 29, 31
 - two-tailed 26, 27, 31
- thalidomide 5–6, 117, 418
- theophylline 49, 391
- therapeutic alternatives 82
- therapeutic ratio 49, 479
- therapeutic risk management *see* risk management
- thiazolidinediones 181–2
- third-party payers 342
- thromboembolic disease 133, 156
- ticrynafen 6
- time sequence of associations 16
- time-stamping microcircuitry 449–50
- time trade-off 350
- tissue partition coefficient 41
- tolmetin 57, 58
- toxic shock syndrome 60
- toxicity information 57–8
- tracheoesophageal fistula 421
- traditional medicines,
 - co-medication 135
- training 447–8
- transdermal drug absorption 43
- transparency in regulatory action 85
- treatment
 - acceptance 368
 - benefits 50–1
 - discontinuing 367
 - harms 50–1
 - identification 186
 - outcomes 50
 - tracking 186
- triazolam 6
- troglitazone 107, 181–2, 187–8, 426
- L-tryptophan 6
- Tylenol 142
- uncertainty
 - about products 92
 - economic analysis 343
 - principle 315–16
 - risk assessment 384
- UnitedHealth Group 184–7, 208, 252
 - claims data 187
 - comparative
 - characteristics 228, 229
 - health plans 183
 - affiliated 252
 - health professional data 185
 - medical claims 185
 - membership data 185
 - new drugs 235
 - pharmacy claims 185, 186
 - research databases 184–5
- upper limit factor 32
- Uppsala Monitoring Center 117
 - Bayesian Confidence Propagation Neural Network 124
 - signal detection system 124

- use of medication 63
 - adverse event reporting rate 131
 - case-control surveillance 141–2
 - chronic 134
 - common uses 81
 - duration 142, 144, 161, 245
 - effectiveness measures 327
 - effects after long intervals 145
 - frequency 141
 - hospital-based evaluation 434–5
 - hospitals 428
 - indications 192
 - inpatient 233, 235
 - multiple 135
 - nondifferential
 - misclassification 147
 - patterns 74
 - regulatory control 85–6
 - stopping in Prescription-Event Monitoring 157, 160
 - timing 142–3
 - see also* drug utilization *entries*
- user fees, drugs 69
- usual care arm of clinical trials 339
- utility measurement 350

- vaccine(s) 7, 131
 - adverse events 411
 - bias 412
 - clinical problems 411
 - confounding factors 393
 - diphtheria-pertussis 261
 - efficacy studies 329
 - epidemiologic studies 411
 - exposure status
 - misclassification 412
 - methodologic issues 411
 - multiple sclerosis relapse 392
 - outcome event rarity 412
 - safety studies 411
 - sample size for studies 412
 - signal detection 412
 - universal exposures 411
- vaccine adverse event reporting system (VAERS) 103, 412, 413
- Vaccine Safety Datalink (VSD) 176, 412
- valdecixib 107
- validation studies 244, 249
- validity 433
 - criterion 346, 474
 - diagnosis information 433
 - drug-induced birth defects 418
 - drug information in hospital 433
 - face 346, 474
 - internal 405
 - medication events 370
 - quality-of-life measurement
 - instruments 346–7
 - research 306
 - see also* data validity
- valproic acid 114, 419
- valsartan 335–6
- variables
 - continuous 32–3, 245
 - misclassification 243
 - predisposing 406
 - ventricular tachycardia 384
 - verapamil 39, 49, 361
 - visual analogue scales 350
 - vital statistics 19
 - Saskatchewan health services
 - database 193
 - volume data 401, 402
 - volume of distribution 475
 - apparent 42
 - voluntariness 304, 479

 - Walker Data Set 200
 - warfarin 48, 49, 291–2
 - Weber effect 112
 - welfare of patients 305
 - white-coat compliance 371
 - winner's curse phenomenon 297
 - World Health Organization (WHO)
 - database 117–18, 123–6
 - drug utilization definition 400, 475
 - Programme for International Drug Monitoring 117–36
 - World Medical Association
 - Declaration of Helsinki (1964) 303

 - Yasmin 156
 - young people, obesity 64

 - zidovudine 57, 60
 - ziprasidone 7
 - zomepirac 6, 58
 - Z-statistics 357