

Pharmacoepidemiology

THIRD
EDITION

Edited by
BRIAN L. STROM

Pharmacoepidemiology, 3rd Edition. Edited by Brian L. Strom
Copyright © 2000 John Wiley & Sons Ltd
ISBNs: 0-471-89925-9 (Hardback); 0-470-84255-5 (Electronic)

PHARMACOEPIDEMIOLOGY

Third Edition

PHARMACOEPIDEMOLOGY

Third Edition

Edited by

BRIAN L. STROM

University of Pennsylvania, USA

JOHN WILEY & SONS, LTD

Chichester • New York • Weinheim • Brisbane • Singapore • Toronto

Copyright © 2000 by John Wiley & Sons Ltd,
Baffins Lane, Chichester,
West Sussex PO19 1UD, England

National 01243 779777
International (+44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on <http://www.wiley.co.uk>
or <http://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, 90 Tottenham Court Road, London 9HE W1P, UK, without the permission in writing of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

WILEY-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

Jacaranda Wiley 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, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

Library of Congress Cataloging-in-Publication Data

Pharmacoepidemiology/[edited by] Brian L. Strom.—3rd ed.

p. ; cm.

Includes bibliographical references and index.

ISBN 0-471-89925-9 (cased : alk. paper)

1. Pharmacoepidemiology. 2. Pharmacology. I. Strom, Brian L.

[DNLM: 1. Pharmacoepidemiology—methods. QZ 42 P536 2000]

RM302.5 P.53 2000

615'.7000004—dc21

99-059591

British Library Cataloguing in Publication Data

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

ISBN 0-471-89925-9

Typeset in 10/12pt Times from the author's disks by Mathematical Composition Setters Ltd, Salisbury, Wiltshire
Printed and bound in Great Britain by Bookcraft (Bath) Ltd, Midsomer Norton
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	xv
Preface to Second Edition	xix
Preface to First Edition	xxi
Acknowledgements	xxiii
PART I. INTRODUCTION	1
1. What is Pharmacoepidemiology?	3
<i>Brian L. Strom</i>	
2. Study Designs Available for Pharmacoepidemiology Studies	17
<i>Brian L. Strom</i>	
3. Sample Size Considerations for Pharmacoepidemiology Studies	31
<i>Brian L. Strom</i>	
4. Basic Principles of Clinical Pharmacology Relevant to Pharmacoepidemiology Studies	41
<i>Patricia McGettigan, David A. Henry and Sean Hennessy</i>	
5. When Should One Perform Pharmacoepidemiology Studies?	63
<i>Brian L. Strom</i>	
PART II. PERSPECTIVES ON PHARMACOEPIDEMIOLOGY	73
6. The Public Health, the University, and Pharmacoepidemiology	75
<i>Paul D. Stolley and Joan-Ramon Laporte</i>	
7. A View From Industry	91
<i>Dale B. Glasser and Gretchen S. Dieck</i>	
8. A View From Regulatory Agencies	109
<i>David J. Graham, Patrick C. Waller and Xavier Kurz</i>	
9. A View From a US Courtroom	125
<i>James R. Ruger and Sidney H. Willig</i>	

PART III. SYSTEMS AVAILABLE FOR PHARMACOEPIDEMOLOGY STUDIES.....	149
10. Spontaneous Reporting in the United States	151
<i>Dianne L. Kennedy, Stephen A. Goldman and Ralph B. Lillie</i>	
11. Spontaneous Reporting Systems Outside the US	175
<i>Bengt-Erik Wiholm, Sten Olsson, Nicholas Moore and Patrick Waller</i>	
12. Intensive Hospital-based Cohort Studies	193
<i>Keith Beard and David H. Lawson</i>	
13. Case–Control Surveillance	209
<i>Samuel Shapiro</i>	
14. Prescription-event Monitoring	231
<i>Ronald D. Mann</i>	
15. Group Health Cooperative of Puget Sound	247
<i>Kathleen W. Saunders, Robert L. Davis and Andy Stergachis</i>	
16. Kaiser Permanente Medical Care Program: Division of Research, Northern California, and Center for Health Research, Northwest Division	263
<i>Gary D. Friedman, Laurel A. Habel, Myde Boles and Bentson McFarland</i>	
17. Harvard Pilgrim Health Care/Harvard Vanguard Medical Associates	285
<i>K. Arnold Chan and Richard Platt</i>	
18. UnitedHealth Group	295
<i>Deborah Shatin, Carol Drinkard and Andy Stergachis</i>	
19. Medicaid Databases	307
<i>Jeffrey L. Carson, Wayne A. Ray and Brian L. Strom</i>	
20. Health Databases in Saskatchewan	325
<i>Winanne Downey, Patricia Beck, Mary McNutt, MaryRose Stang, William Osei and Jim Nichol</i>	
21. Automated Pharmacy Record Linkage in The Netherlands	347
<i>Hubert G. Leufkens and John Urquhart</i>	
22. The Tayside Medicines Monitoring Unit (MEMO)	361
<i>Josie M.M. Evans and Thomas M. MacDonald</i>	
23. The UK General Practice Research Database	375
<i>Luis A. García Rodríguez, Susanne Pérez-Gutthann and Susan Jick</i>	
24. Other Approaches to Pharmacoepidemiology Studies	387
<i>Brian L. Strom</i>	
25. How Should One Perform Pharmacoepidemiology Studies? Choosing Among the Available Alternatives	401
<i>Brian L. Strom</i>	
PART IV. SELECTED SPECIAL APPLICATIONS AND METHODOLOGICAL ISSUES IN PHARMACOEPIDEMOLOGY	415
26. Bioethical Issues in Pharmacoepidemiology Research	417
<i>David Casarett, Jason Karlawish, Elizabeth Andrews and Arthur Caplan</i>	
27. National Medicinal Drug Policies: their Relationship to Pharmacoepidemiology	433
<i>Suzanne Hill and David Henry</i>	
28. Premarketing Applications of Pharmacoepidemiology	449
<i>Harry A. Guess</i>	

29. Studies of Drug Utilization	463
<i>David Lee and Ulf Bergman</i>	
30. Evaluating and Improving Physician Prescribing	483
<i>Stephen B. Soumerai, Sumit Majumdar and Helene L. Lipton</i>	
31. Drug Utilization Review	505
<i>Sean Hennessy, Brian L. Strom, Helene L. Lipton and Stephen B. Soumerai</i>	
32. Determining Causation from Case Reports	525
<i>Judith K. Jones</i>	
33. The Use of Randomized Controlled Trials for Pharmacoepidemiology Studies	539
<i>Samuel M. Lesko and Allen A. Mitchell</i>	
34. The Use of Pharmacoepidemiology to Study Beneficial Drug Effects	553
<i>Brian L. Strom and Kenneth L. Melmon</i>	
35. Pharmacoeconomics: Economic Evaluation of Pharmaceuticals	573
<i>Kevin A. Schulman, Henry Glick, Daniel Polsky and John M. Eisenberg</i>	
36. Using Quality of Life Measurements in Pharmacoepidemiology Research	603
<i>Gordon H. Guyatt and Roman Jaeschke</i>	
37. N-of-1 Randomized Clinical Trials in Pharmacoepidemiology	615
<i>Gordon H. Guyatt, Roman Jaeschke and Robin Roberts</i>	
38. The Use of Meta-analysis in Pharmacoepidemiology	633
<i>Jesse A. Berlin</i>	
39. Validity of Pharmacoepidemiology Drug and Diagnosis Data	661
<i>Suzanne L. West, Brian L. Strom and Charles Poole</i>	
40. Special Methodological Issues in Pharmacoepidemiology Studies of Vaccine Safety	707
<i>Robert T. Chen</i>	
41. Special Methodological Issues in Pharmacoepidemiology Studies of Devices	733
<i>Roselie A. Bright</i>	
42. Special Considerations in Studies of Drug-induced Birth Defects	749
<i>Allen A. Mitchell</i>	
43. Bias and Confounding in Pharmacoepidemiology	765
<i>Jean-Paul Collet and Jean-François Boivin</i>	
44. Novel Approaches to Pharmacoepidemiology Study Design and Statistical Analysis	785
<i>Samy Suissa</i>	
PART V. CONCLUSION	807
45. The Future of Pharmacoepidemiology	809
<i>Brian L. Strom</i>	
Appendix A—Sample Size Tables	817
Appendix B—Glossary	851
Index	859

Contributors

- ELIZABETH ANDREWS, PhD, MPH Director, Worldwide Epidemiology Head, Glaxo Wellcome, Inc., 5 Moore Drive, PO Box 13398, Research Triangle Park, NC 27709, USA, eba43848@glaxowellcome.com
- KEITH BEARD, BSc, MB, FRCP Consultant Physician, South Glasgow University Hospitals NHS Trust Mansionhouse Unit, Victoria Infirmary, Glasgow G41 3DX, UK, keith.beard@btinternet.com
- PATRICIA BECK, MSc Research Consultant, Research Services, Population Health Branch, Saskatchewan Health, 1st Floor East, 3475 Albert Street, Regina, Saskatchewan S4S 6X6, Canada, pbeck@health.gov.sk.ca
- ULF BERGMAN, MD, PhD Associate Professor, Division of Clinical Pharmacology, Karolinska Institute, Huddinge University Hospital, SE-141 86 Stockholm, Sweden, ulf.bergman@pharmlab.hs.sll.se
- JESSE A. BERLIN, ScD Professor of Biostatistics, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 611 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, USA, jberlin@ceeb.upenn.edu
- JEAN-FRANÇOIS BOIVIN, MD, ScD Professor, Department of Epidemiology and Biostatistics, McGill University, 1020 Pine Avenue West, Montreal, Quebec H3A 1A2, Canada, jeanfb@epid.lan.mcgill.ca
- MYDE BOLES, PhD Investigator, Center for Health Research, Kaiser Permanente Northwest, 3800 N Interstate Avenue, Portland, OR 97227, USA, myde.boles@kp.org
- ROSELIE A. BRIGHT, ScD Epidemiologist, Epidemiology Branch, Division of Postmarket Surveillance, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, US Food and Drug Administration, 1350 Piccard Drive, HFZ-541, Rockville, MD 20850, USA, rxb@cdrh.fda.gov
- ARTHUR CAPLAN, PhD Director, Center for Bioethics, University of Pennsylvania, 3401 Market Street, Suite 320, Philadelphia, PA 19104-3308, USA, caplan@mail.med.upenn.edu
- JEFFREY L. CARSON, M.D. Richard C. Reynolds Professor of Medicine, Chief, Division of General Internal Medicine, UMDNJ-Robert Wood Johnson Medical School,

- Room 2302A, Clinical Academic Building, 125 Paterson Street, New Brunswick, NJ 08903-0019, USA, carson@umdnj.edu
- DAVID CASARETT, MD, MA
Instructor, Division of Geriatrics, University of Pennsylvania, Institute on Aging, 3615 Chestnut Street, Philadelphia, PA 19104, USA, casarett@mail.med.upenn.edu
- K. ARNOLD CHAN, MD, DS
Assistant Professor of Epidemiology, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA, kachan@epinet.harvard.edu
- ROBERT T. CHEN, MD, MA
Chief, Vaccine Safety and Development Branch (E61), National Immunization Program, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, GA 30333, USA, bchen@cdc.gov
- JEAN-PAUL COLLET, MD, PhD, MSc
Associate Professor, McGill University, Director, Randomized Clinical Trial Unit, Jewish General Hospital, 3755 Cote-Ste-Catherine Road, Montreal, Quebec H3T IE2, Canada, jpcollet@epid.jgh.mcgill.ca
- ROBERT L. DAVIS, MD, MPH
Associate Professor, Departments of Pediatrics & Epidemiology, Box 358853, University of Washington Medical Center, Seattle, WA 98195 USA, rdavis@u.washington.edu
- GRETCHEN S. DIECK, PhD
Senior Director, Worldwide Safety, Pfizer Inc, 235 East 42nd Street, 7th Floor, New York, NY 10017, USA, dieckg1@pfizer.com
- WINANNE DOWNEY, BSP
Manager, Research Services, Population Health Branch, Saskatchewan Health, 1st Floor East, 3475 Albert Street, Regina, Saskatchewan S4S 6X6, Canada, wdowney@health.gov.sk.ca
- CAROL R. DRINKARD, PhD
Senior Research Analyst, Center for Health Care Policy and Evaluation, UnitedHealth Group, P.O. Box 1459, Minneapolis, MN 55440-1459, USA, carol_r_drinkard@uhc.com
- JOHN M. EISENBERG, MD, MBA
Director, Agency for Healthcare Research and Quality, Department of Health and Human Services, 2101 East Jefferson Street, Suite 600, Rockville, MD 20852, USA, jeisenbe@ahrq.gov
- JOSIE M. M. EVANS, MA (Oxon) PhD
Wellcome Research Fellow, Medicines Monitoring Unit, Department of Clinical Pharmacology, University of Dundee, Ninewells Hospital and Medical School, Dundee DDI 9SY, UK, josie@memo.dundee.ac.uk
- GARY D. FRIEDMAN, MD, MS
Adjunct Investigator, Division of Research, Kaiser Permanente Medical Care Program, Northern California, 3505 Broadway, Oakland, CA 94611, USA, Consulting Professor, Stanford University School of Medicine, Department of Health Research and Policy, Redwood Building, Room T210, Stanford, CA 94305-5405, USA, gdf@dor.kaiser.org
- DALE B. GLASSER, PhD
Senior Associate Director/Group Leader, Safety Evaluation and Epidemiology, Pfizer Inc, 235 East 42nd Street, MS 150-3-75, New York, NY 10017, USA, dale.glasser@pfizer.com
- HENRY GLICK, PhD
Assistant Professor, General Internal Medicine, University of Pennsylvania School of Medicine, 1211 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, USA, hlthsvrs@mail.med.upenn.edu
- STEPHEN A. GOLDMAN, MD
Former Medical Director, MedWatch, Food and Drug Administration; Director, Pharmacoepidemiology, Knoll Pharmaceutical Company, 3000

- Continental Drive-North, Mount Olive, NJ 07828-1234, USA, goldmas@basf.com
- DAVID J. GRAHAM, MD, MPH Associate Director for Science and Medicine, Office of Postmarketing Drug Risk Assessment, Center for Drug Evaluation and Research, Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, USA grahamd@cder.fda.gov
- HARRY A. GUESS, MD, PhD Executive Director, Epidemiology Department, Merck Research Laboratories BL-1-7, PO Box 4, West Point, PA 19486-0004, USA, harry_guess@merck.com
- GORDON H. GUYATT, MD, FRCP(C) Professor of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Center, 1200 Main Street West, Room 2C12, Hamilton, Ontario L8N 3Z5, Canada, guyatt@fhs.csu.mcmaster.ca
- LAUREL A. HABEL, PhD Research Scientist, Division of Research, Kaiser Permanente Medical Care Program, Northern California, 3505 Broadway, Oakland, CA 94611, USA, lah@dor.kaiser.org
- SEAN HENNESSY, PharmD, MSCE Senior Research Investigator, Center for Clinical Epidemiology and Epidemiological Biostatistics, University of Pennsylvania School of Medicine, 803 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, USA, shenness@cceb.med.upenn.edu
- DAVID A. HENRY MBChB, FRCP Professor, Discipline of Clinical Pharmacology, Faculty of Medicine and Health Sciences, University of Newcastle, Newcastle Mater Hospital, Waratah NSW 2298, Australia, mddah@alinga.newcastle.edu.au
- SUZANNE HILL, MBBS, PhD, FAFPHM Senior Lecturer, Discipline of Clinical Pharmacology, School of Population Health Sciences, Faculty of Medicine and Health Sciences, University of Newcastle, New South Wales, Australia 2308, hillsu@mail.newcastle.edu.au
- ROMAN JAESCHKE, MD, MSc Associate Clinical Professor of Medicine, Department of Medicine, St. Joseph's Hospital, 301 James Street S., Hamilton, Ontario L8N 3A6, Canada, jaeschke@fhs.csu.mcmaster.ca
- SUSAN JICK, DSc Associate Professor of Epidemiology and Biostatistics, Boston University School of Public Health, Co-director, Boston Collaborative Drug Surveillance Program, 11 Muzzey Street, Lexington, MA 02421-5207, USA, sjick@bu.edu
- JUDITH K. JONES, MD, PhD President, The Degge Group, Ltd., Suite 1430, 1616 N. Ft. Myer Drive, Arlington, VA 22209-3109, USA, jkjoness@deggegroup.com
- JASON KARLAWISH, MD Assistant Professor of Medicine, 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
- DIANNE L. KENNEDY, MPH, RPh Former Director, MedWatch, Food and Drug Administration, Project Manager, Food and Drug Administration Pregnancy Labeling Task Force, HFD-104, Center for Drug Evaluation and Research, Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, USA, kennedyd@cder.fda.gov
- XAVIER KURZ, MD, MSc, PhD Scientific Coordinator, Belgian Centre for Pharmacovigilance, Ministry of Public Health, Brussels, Belgium, xavier.kurz@afigp.fgov.be

- JOAN-RAMON LAPORTE, MD
Director, Fundació Institut Català de Farmacologia, Universitat Autònoma de Barcelona, Hospitals Vall d'Hebron, 08035-Barcelona, Spain, jrl@icf.uab.es
- DAVID H. LAWSON, CEB, MD, FRCP
Consultant Physician, Royal Infirmary, Glasgow G4 OSF, UK, ann.rodde@northglasgow.scot.nhs.uk
- DAVID LEE, MD
Deputy Director, Drug Management Program, Management Sciences for Health, 1515 Wilson Boulevard, Suite 710, Arlington, VA 22209, USA, dlee@msh.org
- SAMUEL M. LESKO, MD, MPH
Senior Epidemiologist, Slone Epidemiology Unit, Associate Professor of Public Health, Boston University Schools of Public Health and Medicine, 1371 Beacon Street, 3rd Floor, Brookline, MA 02446, USA, slesko@slone.bu.edu
- HUBERT G. M. LEUFKENS, PhD
Chair, Department of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute for Pharmaceutical Sciences (UIPS), PO Box 80082, 3508 TB Utrecht, The Netherlands, h.g.m.leufkens@pharm.uu.nl
- RALPH B. LILLIE, MPH, RPh
Former Deputy Director, Office of Postmarketing Drug Risk Assessment, Center for Drug Evaluation and Research, Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, USA, lillie@ceder.fda.gov
- HELENE LEVENS LIPTON, PhD
Professor of Pharmacy and 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 265, San Francisco, CA 94118, USA, lipton@itsa.ucsf.edu
- SUMIT MAJUMDAR, MD, MPH
Assistant Professor, Division of General Internal Medicine, University of Alberta, Edmonton, Alberta T6G 2B7, Canada, me2.majumdar@ualberta.ca
- RONALD D. MANN, MD
Emeritus Professor, Drug Safety Research Unit, Southampton, UK. Send correspondence to Dr Mann at: Manor Cottage, Petersfield Road, Midhurst, West Sussex GU29 9RL, UK, dr.mann@manorcottage.fsbusiness.co.uk
- THOMAS M. MACDONALD, MD
Professor of Clinical Pharmacology and Pharmacoepidemiology, Director, Medicines Monitoring Unit (MEMO), Department of Clinical Pharmacology and Therapeutics, University of Dundee, Ninewells Hospital and Medical School, Dundee DDI 9SY, UK, tom@memo.dundee.ac.uk
- BENTSON H. MCFARLAND, MD, PhD
Adjunct Investigator, Kaiser Permanente Center for Health Research, Professor of Psychiatry, Public Health and Preventive Medicine, Oregon Health Sciences University, Professor of Pharmacy, Oregon State University College of Pharmacy, 3800 North Interstate Ave., Portland, OR 97227, USA, mcfarlandbe@chr.mts.kpnw.org
- PATRICIA MCGETTIGAN, B Pharm, MD, MRCP, FRACP
Research Academic, Discipline of Clinical Pharmacology, School of Population Health Sciences, Faculty of Medicine and Health Sciences, University of Newcastle, Mater Misericordiae Hospital, Waratah, NSW 2298, Australia, gmgetti@mail.newcastle.edu.au
- MARY MCNUTT, MSc
Research Consultant, Research Services, Population Health Branch, Saskatchewan Health, 1st Floor East, 3475 Albert Street, Regina, Saskatchewan S4S 6X6, Canada, mmcnett@health.gov.sk.ca
- KENNETH L. MELMON, MD
Professor of Medicine and Molecular Pharmacology, Associate Dean of Post Graduate Medical Education, 341 Medical School Office Building, Stanford University School of Medicine, Stanford, CA 94305, USA, melmon@smi.stanford.edu

- ALLEN A. MITCHELL, MD
Director, Slone Epidemiology Unit, Professor of Public Health and Pediatrics, Boston University Schools of Public Health and Medicine, 1371 Beacon Street, Brookline, MA 02446, USA, amitchell@slone.bu.edu
- NICHOLAS MOORE, MD, PhD
Professor of Clinical Pharmacology, Department of Pharmacology, Victor Segalen University, CHU Pellegrin Carreire, 33076 Bordeaux Cédex, France, nicholas.moore@pharmaco.u-bordeaux2.fr
- JIM NICHOL, BAdmin
Research Consultant, Research Services, Population Health Branch, Saskatchewan Health, 1st Floor East, 3475 Albert Street, Regina, Saskatchewan S4S 6X6, Canada, jnichol@health.gov.sk.ca
- STEN OLSSON, MSci Pharm
Head, External Affairs, The Uppsala Monitoring Center, Stora Torget 3, S-573 20 Uppsala, Sweden, sten.olsson@who-umc.org
- WILLIAM OSEI, MBChB, MPH
Director, Epidemiology, Research and Evaluation Unit, Population Health Branch, Saskatchewan Health, 1st Floor East, 3475 Albert Street, Regina, Saskatchewan S4S 6X6, Canada, wosei@health.gov.sk.ca
- SUSANNE PEREZ-GUTTHANN, MD, MPH
Head, Global Medical Safety and Global Epidemiology, Novartis, Gran Via Corts Catalanes, 764, 08013 Barcelona, Spain, susana.perez@pharma.novartis.com
- RICHARD PLATT, MD, MSc
Professor of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 126 Brookline Ave, Suite 200, Boston, MA 02215, USA, richard.platt@channing.harvard.edu
- DANIEL POLSKY, PhD
Research Assistant Professor, 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 7400), School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400, USA, cpoole@unc.edu
- WAYNE A. RAY, PhD
Professor and Director, Division of Pharmacoepidemiology, Vanderbilt University School of Medicine, Medical Center North A-1106, Nashville, TN 37232-2637, USA, wayne.ray@mcmail.vanderbilt.edu
- ROBIN ROBERTS, MSc
Professor, Clinical Epidemiology and Biostatistics, Hamilton Civic Hospitals Research Center, Henderson General Division, 711 Concession Street, Hamilton, Ontario L8V1C3, Canada, robertsr@mcmaster.ca
- LUIS ALBERTO GARCÍA RODRÍGUEZ, MD, MSc
CEIFE – Centro Español de Investigación Farmacoepidemiológica, Almirante, 28-2º, Madrid 28004, Spain, lagarcia@ceife.es
- JAMES R. RUGER, Ph.D, Esq
Associate General Counsel, Siemens Corporation, 186 Wood Avenue South, Iselin, NJ 08830, USA, james.ruger@sms.siemens.com
- KATHLEEN W. SAUNDERS, JD
CHS Analyst/Programmer III, The Center for Health Studies, Group Health Cooperative of Puget Sound, 1730 Minor Ave, Suite 1600, Seattle, WA 98101, USA, saunders.k@ghc.org
- KEVIN A. SCHULMAN, MD
Associate Professor of Medicine, Director, Center for Clinical and Genetic Economics, Duke Clinical Research Institute, PO Box 17969, Durham, NC 22715, USA, schulO12@mc.duke.edu

- SAMUEL SHAPIRO, MB, FRCP (E) Director Emeritus, Slone Epidemiology Unit, Boston University Schools of Public Health and Medicine, Professor of Epidemiology, Columbia University, School of Public Health, 622 West 168th Street, PH18-107, New York, NY 10032, USA, sydzeev@aol.com
- DEBORAH SHATIN, PhD Senior Researcher, Center for Health Care Policy and Evaluation, UnitedHealth Group, P.O. Box 1459, Minneapolis, MN 55440-1459, USA, deborah_shatin@uhc.com
- STEPHEN B. SOUMERAI, ScD Professor of Ambulatory Care and Prevention, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 126 Brookline Avenue, Suite 200, Boston, MA 02215, USA, ssoumerai@hms.harvard.edu
- MARYROSE STANG, PhD Research Consultant, Research Services, Population Health Branch, Saskatchewan Health, 1st Floor East, 3475 Albert Street, Regina, Saskatchewan, S4S 6X6, Canada mstang@health.gov.sk.ca
- ANDY STERGACHIS, PhD, RPh Director of Pharmacy and Quality Assurance, Drugstore.com, 18650 NE 67th Court, Bellevue, WA 98052, USA, andy@drugstore.com
- PAUL D. STOLLEY, MD, MPH Professor of Epidemiology, Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, 660 West Redwood Street, Suite 109, Baltimore, MD 21201, USA, pstolley@epi.umaryland.edu
- BRIAN L. STROM, MD, MPH Professor of Biostatistics and Epidemiology, Professor of Medicine, Professor of Pharmacology, University of Pennsylvania School of Medicine, Chair, Department of Biostatistics and Epidemiology, Director, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 824 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19014-6021, USA, bstrom@cceb.med.upenn.edu
- SAMY SUISSA, PhD Professor of Epidemiology and Biostatistics, Professor of Medicine, McGill University, Head, McGill Pharmacoepidemiology Research Unit, Director, Division of Clinical Epidemiology, Royal Victoria Hospital, 687 Pine Avenue West, Ross 4, Montreal PQ H3A 1A1, Canada, samy.suissa@clinepi.mcgill.ca
- JOHN URQUHART, MD, FRCP(Edin) Professor of Pharmaco-epidemiology, Maastricht University, Maastricht, Netherlands, Chief Scientist, AARDEX Ltd/APREX Corp, Zug, Switzerland & Union City, CA USA, Professor of Biopharmaceutical Sciences, UCSF, San Francisco, CA USA, Send correspondence to: 975 Hamilton Ave, Palo Alto, CA 94301, USA, urquhart@ix.netcom.com
- PATRICK C. WALLER, MD, FRCP (Ed), FFPM, MPH Specialist in Pharmacovigilance, Post-Licensing Division, Medicines Control Agency, Market Towers, I Nine Elms Lane, London SW8 5NQ, UK, patrick.waller@mca.gov.uk
- SUZANNE L. WEST, PhD, MPH Director of Research, The North Carolina Program for Women's Research, University of North Carolina at Chapel Hill, Cecil G. Sheps Center for Health Services Research, 725 Airport Road CB#7590, Chapel Hill, NC 27599-7590, USA, sue_west@med.unc.edu
- BENGT-ERIK WIHOLM, MD, PhD, FRCP(L) Senior Director, Epidemiology, Global Drug Surveillance and Outcome Research, Pharmacia Corporation, 100 Route 206 North, Peapack, NJ 07977, USA, bengt-erik.wiholm@am.pnu.com
- SIDNEY H. WILLIG, JD Professor Emeritus, Temple University, School of Law, 1992 N. Broad Street, Philadelphia, PA 19122, USA

Preface

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

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

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

The 1970s, 1980s, and 1990s continued to see a series of accusations about major adverse events possibly associated with drugs, ranging from those discussed in the first edition of this book (liver disease caused by benoxaprofen, subacute myeloptico-neuropathy (SMON) caused by clioquinol, the oculomucocutaneous syndrome caused by practolol, acute flank pain and renal failure caused by suprofen, liver disease caused by ticrynafen,

anaphylactoid reactions caused by zomepirac), to those added in the second edition (arrhythmias from astemizole; hypertension, seizures, and strokes from postpartum use of bromocriptine; deaths from fenoterol; suicidal ideation from fluoxetine; hypoglycemia from human insulin; birth defects from isotretinoin; cancer from depot-medroxyprogesterone; multiple illnesses from silicone breast implants; memory and other central nervous system disturbances from triazolam; arrhythmias from terfenadine; hemolytic anemia and other adverse reactions from temafloxacin), to the new list added in this edition (including liver toxicity from amoxicillin-clavulanic acid; liver toxicity from bromfenac; cancer and myocardial infarction 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; eosinophilia myalgia from tryptophan; anaphylactic reactions from vitamin K; and liver toxicity from troglitazone). Recently published data also suggest that adverse drug reactions could be as much as the fourth leading cause of death. These and other serious but uncommon drug effects have led to the development of new methods to study drug effects in large numbers of patients. Academic investigators, the pharmaceutical industry, regulatory agencies, and the legal

community have turned for these methods to the field of epidemiology, the study of the distribution and determinants of disease in populations.

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

The field of pharmacoepidemiology has grown enormously since the publication of the first edition of this book. The International Society of Pharmacoepidemiology, simply an early idea when the first edition of this book was written, has grown to a major international scientific force, with over 1200 members from 57 countries, an extremely successful and well attended annual meeting, a large number of very active committees, and its own journal. At least four other journals have been founded as well, all competing to publish the work of this field, and a number of established journals have targeted pharmacoepidemiology manuscripts as desirable. As new scientific developments occur within mainstream epidemiology, they are rapidly adopted, applied, and advanced within our field. 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, pharmacoconomics, and quality-of-life studies. The continuing parade of drug safety crises continues to emphasize the need for the field, and some foresighted manufacturers have begun to perform prophylactic pharmacoepidemiology

studies, in order to have data in hand and available when questions occur, rather than waiting until a crisis has developed. Pharmacoepidemiology data are now routinely used for regulatory decisions, and many governmental agencies have been developing and expanding their own pharmacoepidemiology programs. 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 in response to a crying need for more manpower, a few of them have also established pharmacoepidemiology training programs. Pharmacoepidemiology research funding is now more plentiful, and even limited support for training is now available. An international foundation was formed, with one of its missions to support pharmacoepidemiology work, and then disbanded.

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. 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. As this book goes to press, the United States is again debating whether Medicare should begin providing drug benefits, that is, paying for drugs for those over age 65. If it does, this should generate an enormous new interest in this field, as a major branch of the federal government becomes concerned about the costs of 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 pharmacoepidemiology research much more difficult.

In the first edition, my goal was to help to introduce this new field to the scientific world. The

explosion in interest in the field, the rapid scientific progress that has been made, and the unexpected sales of the first edition led to the second edition. The continued maturation of what used to be a new field, the marked increase in sales of the second edition over the first, and the many requests I have had from people all over the world have now led me to organize this third edition. Much in the field has changed, and so much of the book has changed. Most chapters have been thoroughly revised. A number of new chapters have been added, and many new authors. Overall, the book has continued to expand in size, although with some careful pruning of old chapters, the net growth has been kept to only four chapters.

As in earlier editions, Part I of this book provides background information on what is included in the field of Pharmacoepidemiology, a description of the epidemiologic study designs it uses, a description of its unique problem—the requirement for very large sample sizes—and a discussion about when one would want to perform a pharmacoepidemiology study. Also included is a chapter providing analogous basic principles of clinical pharmacology. Part II presents a series of discussions on the need for the field, the contributions it can make, and some of its problems, from the perspectives of the Food and Drug Administration, the pharmaceutical industry, academia, and the legal community. Part III describes the systems that have been developed to perform pharmacoepidemiologic studies, and how each approaches the problem of gathering large sample sizes of study subjects in a cost-effective manner. A number of new data resources have been developed, others lost, and some lost and then revived, with changes occurring even as we go to press. Part IV presents state-of-the-art discussions of

some particular methodologic issues and opportunities that have arisen in the field. These are of particular interest as the field continues to turn its attention to questions beyond just those of adverse drug reactions. Finally, Part V provides my personal speculations as to the future of the field. My expectation is that Parts I, II, III, and V of this book will be of greatest interest to the neophyte. In contrast, Parts II, III, IV, and V should be of greatest interest to those with some background, who want a more in-depth view of the field.

This book is not intended as a textbook of adverse drug reactions, that is a compilation of drug-induced problems organized either by drug or by problem. Several of these already exist. Rather, it is intended to elucidate the methods of investigating adverse drug reactions, as well as other questions of drug effects. It is also not intended as a textbook of clinical pharmacology, organized by disease or by drug, or of epidemiology, but a text describing the overlap between the two fields.

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

Brian L. Strom, MD, MPH, 1999

Preface to the Second Edition

The field of pharmacoepidemiology has grown enormously since the publication of the first edition of this book. The International Society of Pharmacoepidemiology has grown from an idea held by a few diehard stalwarts to a major international scientific force, with over 900 members from 40 countries, an extremely successful and widely attended annual meeting, a large number of very active committees, and its own journal. At least four other journals have been founded as well, all competing to publish the work of this field, and a number of established journals have targeted pharmacoepidemiology manuscripts as desirable. As new scientific developments occur within mainstream epidemiology, they are rapidly adopted, applied, and advanced within our field. 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 new Clinical Pharmacology board examination.

Most of the major international pharmaceutical companies have founded special units to organize and lead their efforts in pharmacoepidemiology, pharmacoconomics, and quality-of-life studies. Pharmacoepidemiology data are now routinely used for regulatory decisions, and many governmental agencies have been developing and expanding their own pharmacoepidemiology programs. Requirements that a drug be proven to be cost effective have been added to many national 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 in response to a crying need for more pharmacoepidemiologists, a few of them have also established training programs. Pharmacoepidemiology research funding is now more plentiful, and even limited support for training is now available. An international foundation has been formed, and one of its missions is to support pharmacoepidemiology work.

In the US, drug utilization review programs are now being required, by law, of each of the 50 state Medicaid programs, and is being discussed as a nationwide requirement as part of national health reform. In addition, the Joint Commission on Accreditation of Health Care Organizations now requires that every hospital in the country have an adverse drug reaction monitoring program and a drug use evaluation program, turning every hospital into a mini-pharmacoepidemiology laboratory. Stimulated in part by the interests of the World Health Organization and the Rockefeller Foundation, there is even substantial interest in pharmacoepidemiology now in the developing world.

In the meantime, the parade of pharmacoepidemiology crises continues, including birth defects from isotretinoin; multiple illnesses from silicone breast implants; memory and other central nervous system disturbances from triazolam; hypoglycemia from human insulin; suicidal ideation from fluoxetine; deaths from fenoterol and maybe other beta agonist inhalers; cancer from depot-medroxyprogesterone; arrhythmias from terfenadine and

astemizole; hypertension, seizures, and strokes from postpartum use of bromocriptine; multiple adverse reactions from temafloxacin; etc. These continue to emphasize the need for the field, and a number of manufacturers have begun to perform prophylactic pharmacoepidemiology studies, in order to have data in hand and available when questions occur, rather than waiting until a crisis has developed.

In the prior edition, my goal was to help to introduce this new field to the scientific world. The explosion in interest in the field, the rapid scientific progress that has been made, the remarkable and unexpected sales of the first edition, and the many requests I have had from people all over the world, have now led me to organize this second edition. Much in the field has changed, and so much of the book has changed. As before, this book is not intended as a textbook of adverse drug reactions; several of these already exist. Rather, it is intended to elucidate the methods of investigating adverse drug reactions, as well as other questions of drug effects. It is also not intended as a textbook of clinical pharmacology, or of epidemiology, but a text describing the overlap between the two fields. As in the previous edition, Part I represents an introduction to the field, and a presentation of basic scientific principles of clinical epidemiology necessary to be a pharmacoepidemiologist. In the current edition, a chapter has now been added, providing analogous basic principles of clinical pharmacology, thereby correcting an omission. Part II remains a presentation of multiple different perspectives on the field. A number of new data resources have been developed, others lost, and some lost and then revived, with changes occurring even as we go to press. This has led to a number of new chapters and some substitutions of chapters in Part III. There have been many major new methodological developments in the field, as well,

including the further development of drug utilization review, meta-analysis, *N*-of-1 clinical trials, and studies of quality-of-life, leading to the addition of six new chapters to Part IV. In addition, five new chapters have been added to Part IV presenting state-of-the-art discussions of selected methodological issues of special importance to pharmacoepidemiology research, including the validity of measurements of drug exposure and disease outcome in pharmacoepidemiology research, study design issues particular to pharmacoepidemiology, biostatistical issues particular to pharmacoepidemiology, etc. In total, 17 of the 41 chapters in this edition are totally new, and many of the remainder have been thoroughly revised. Many new authors have been added, as well. Fortunately or unfortunately, in the process the book has substantially grown, as the field has grown.

In the previous edition, I had hoped that this book could serve as both a reference source and a textbook, and it appears that it indeed served both roles. My hope is that this will continue. As before, Parts I, II, and III of this book will be of greatest interest to the neophyte. In contrast, Parts II, III, and IV should be of greatest interest to those with some background who want a more in-depth view of the field.

I have been excited by the rapid progress and growth our field has seen, and delighted that this book has played a small role in assisting this. With this new edition, it will document the major changes the field has seen. In the process, my hope is that it can continue to serve to assist the field in its development.

Brian L. Strom, MD, MPH
January 1994

Preface to the First Edition

If the whole *materia medica*, as now used, could be sunk to the bottom of the sea, it would be all the better for mankind, and all the worse for the fishes.

Oliver Wendell Holmes
Medical Essays, "Comments and Counter"
Currents in Medical Science

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

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

The 1970s and 1980s have also seen a series of major adverse drug reactions, from subacute myelo-optic-neuropathy (SMON) caused by clioquinol, to the oculomucocutaneous syndrome caused by practolol, to deaths from liver disease caused by ticrynafen, to deaths from liver disease

caused by benoxaprofen, to anaphylactoid reactions caused by zomepirac, and to acute flank pain and renal failure caused by suprofen. These and other serious but uncommon drug effects have led to the development of new methods to study drug effects in large numbers of patients. Academic investigators, the pharmaceutical industry, the Food and Drug Administration, and the legal community have turned for these methods to the field of epidemiology, the study of the distribution and determinants of disease in populations.

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

A new generation of pharmacoepidemiologists is now arising, much larger than the previous one, to join the field's few pioneers. The growing interest in this field is demonstrated by the increasing attendance at the annual International Conference on Pharmacoepidemiology, which has been held each year in Minneapolis. It is also demonstrated by the recent development of two new journals, each claiming to have pharmacoepidemiology as a major focus. There are even tentative plans for the formation of a society of pharmacoepidemiologists.

A recent paper in a pharmacy journal called for instruction in pharmacoepidemiology for *all* doctor of pharmacy students. In the US, Medicare is about to begin providing drug benefits, that is, paying for the drugs for those over age 65. This should generate an enormous new interest in this field, as a major branch of the federal government becomes concerned about the costs of and effects of prescription drugs. In addition, there is discussion in the US of replacing the largest premarketing clinical trials of drugs (“phase 3” testing) with increased postmarketing surveillance, at least for selected drugs. This, too, will lead to a marked increase in interest in the field.

My purpose in writing this book is to document the current state of this new and growing discipline. Although pharmacoepidemiology studies have already made contributions of public health importance, the field has the potential to accomplish much more. However, it is difficult for anyone, even with experience in the field, to gain an overall perspective on its current status. There is no introductory text now available for the neophyte. I am regularly asked to recommend a book for someone who wants to learn about this field, but none has existed.

This book is not intended as a textbook of adverse drug reactions, that is a compilation of drug-induced problems organized either by drug or by problem. Several of these already exist. Rather, it is intended to elucidate the methods of investigating adverse drug reactions, as well as other questions of drug effects. It is also not intended as a textbook of clinical pharmacology, organized by disease or by drug, or of epidemiology, but a text describing the overlap between the two fields.

Specifically, Section I of this book will provide background information on what is included in the field of pharmacoepidemiology, a description of the epidemiologic study designs it uses, a description of its unique problem—the requirement for very large sample sizes, and a discussion about when one would want to perform a pharmacoepidemiology study. Section II will present a series of discussions on the need for the field, the contributions it can make, and some of its problems, from the perspectives of the Food and Drug Administration, the pharmaceutical industry, academia, and the legal community. Section III will describe the systems that have been developed to perform pharmacoepidemiologic studies, and how each approaches the problem of gathering large sample sizes of study subjects in a cost-effective manner. Lastly, Section IV presents some particular methodological issues and opportunities which have recently arisen in the field. This is of particular interest as the field turns its attention to questions beyond just those of adverse drug reactions. My expectation is that Sections I, II, and III of this book will be of greatest interest to the neophyte. In contrast, Sections II, III, and IV should be of greatest interest to those with some background who want a more in-depth view of the field.

It is my hope that this book can serve both as a useful introduction to pharmacoepidemiology and a reference source for the growing number of people interested in this field—in academia, in regulatory agencies, in industry, and in the law. It will also hopefully be useful as a text for the growing number of courses being developed in this field.

Brian L. Strom, MD, MPH, 1986

Acknowledgements

There are many individuals and institutions to whom I owe thanks for their contributions to my efforts in preparing this book. Over the years, my pharmacoepidemiology work has been supported by cooperative agreement FD-U-000079 from the US Food and Drug Administration; NIH grants and contracts R01-HL 27433, R01-AM/HD 31865, 1-R01-AM32869, R01-HD24316, R01-HD21726, R01-HD20531, R01-HL39000, R01-HD29201, R01 AG14601, NICHD CRE-91-3, and PO1 CA77596; grants from the Joint Commission on Prescription Drug Use, the Asia Foundation, the Charles A. Dana Foundation, The Rockefeller Foundation, the Andrew W. Mellon Foundation, and the International Clinical Epidemiology Network, Inc.; and grants from Alza Corporation, Bayer Corporation, Berlex Laboratories, the Burroughs Wellcome Company, Ciba-Geigy Corporation, Health Information Designs, Inc., Hoechst-Roussel Pharmaceuticals, Hoffman-La Roche, Inc., International Formula Council, Marion Merrell Dow, Inc., McNeil Consumer Products, McNeil Pharmaceuticals, Mead Johnson Pharmaceuticals, Merck and Company, Novartis Pharmaceuticals Corp., Pfizer Pharmaceuticals, A. H. Robins Company, Rowell Laboratories, Sandoz Pharmaceuticals, Schering Corporation, Smith Kline and French Laboratories, Sterling Winthrop Inc., Syntex, Inc., the Upjohn Company, and Wyeth-Ayerst Research. In addition, generous support to our pharmacoepidemiology training program has been provided by Alza Corporation, Berlex Laboratories, Inc.,

Ciba-Geigy Corporation, Hoechst-Marion-Roussel, Inc., Merck and Company, Inc., McNeil Consumer Products, Novartis Pharmaceuticals Corporation, Pfizer Pharmaceuticals, Integrated Therapeutics, Inc., a subsidiary of Schering-Plough Corporation, SmithKline Beecham Pharmaceuticals, and Wyeth-Ayerst Pharmaceuticals. While none of this support was specifically intended to support the development of this book, without this assistance I would not have been able to support my career in pharmacoepidemiology.

Jane Sinopoli-Sosa, Barbara Price, and Madelaine Sauk were their usual dedicated and hard-working selves, assisting me in the myriad arrangements which the book required. Rita Schinnar and Judith Kinman helped proof the galleys. Finally, Anne Saint John provided superb help in preparing both the manuscripts for my chapters and the edited versions of most of the other chapters.

I would like to thank my parents for the support and education which were critical to my being able to begin my career. I would also like to thank Paul D. Stolley, M.D., M.P.H. and Kenneth L. Melmon, M.D. for their direction, guidance, and inspiration in the formative years of my career. I would like to thank my trainees, from whom I learn at least as much as I teach. Last, but certainly not least, I would like to thank my family—Lani, Shayna, and Jordi—for accepting the time demands of the book, for tolerating my endless hours working at home on it, and for their ever present love and support.

Index

Note: page references in *italics* refer to figures; those in **bold** refer to tables

- absorption, variation in, 49–50
- Abt Associates Inc., 516
- acarbose, 242
- accuracy, 765
- Accutane, 7, 164, 751, 753
- acetaminophen, 55, 222, 542
- acetazolamide, 187
- active tubular secretion, 51
- Activities of Daily Living, 268
- acute flank pain syndrome, 188–9
- acyclovir, 256, 278
- adaptive responses, 44–5
- ADE Prevention Study Group, 196
- Adverse and Allergic Drug Event Reporting, 268
- Adverse Drug Reaction System (FDA), 713
- adverse drug reactions (ADRs)
 - historical background, 5–8
 - incidence, 3
 - reporting, 109
 - see also* Adverse Event Reporting System
- adverse effect studies, 6, 7
- Adverse Event Reporting System (AERS) (FDA), 152, 153, 155, 162–3, 168–9
- adverse event studies, 6, 7
- Adverse Events Following Immunizations (AEFI), 715
- Adverse Reaction Identification Evaluation System (ARIES), 202
- Advice Nurse Documentation System (RANDS), 268
- age, effects of, 45
- Agency for Health Care Policy and Research (AHCPR), 97, 420, 743, 813, 815
 - CERTT program, 813
- Agency for Healthcare Research and Quality (AHRQ), *see* Agency for Health Care Policy and Research
- Agent Orange, 130, 131, 132, 144
 - litigation, 133, 134, 145
- agranulocytosis, 23, 37, 184, 186, 223, 394
- Aid to Families with Dependent Children Program (AFDC), 307
- AIDS, 725
- alcohol/breast cancer relationship, 215, 222
- alendronate, 242, 301
- alkylating agents, 96
- Allen v. The United States*, 132
- allopurinol, 201
- alloyed gold standard, 664
- alpha error, 32
- alprazolam, 118, 280
- alprenolol, 54
- American Association of Retired Persons (AARP), 296
- American Hospital Association (AHA)
 - Annual Survey of Hospitals, 743
- American Hospital Formulary System (AHFS), 286, 473
- American Medical Association
 - Committee on Blood Dyscrasias, 153
 - Council on Pharmacy and Chemistry, 6
- American Rheumatism Association, 607
- American Society for Clinical Pharmacology and Therapeutics, 8
- aminoglycoside, 198
- aminophylline, 4
- amitriptyline/amitriptyline, 54
- amoxicillin, 303, 380–1
- amoxicillin–clavulanic acid, 7, 117
- amphetamine, 199
- ampicillin, 186
- analyses of secular trends, 23–4
 - data sources, 387–91
- analysis of variance (ANOVA), 591, 592, 621
- analytic studies, 27
- Anatomic Therapeutic Chemical (ATC) classification system, 472
- Andrew W. Mellon Foundation, 814
- angiotensin converting enzyme, 43
- angiotensin converting enzyme inhibitors (ACEIs), 52, 180, 381, 540
- angiotensin I, 43
- angiotensin II, 43, 46
- antibiotics
 - birth defects and, 301–2
 - diarrhea and, 302–3
 - liver injury and, 380–1
- anticipated beneficial effects, 554
- anticipated harmful effects, 553–4
- anticoagulants
 - myocardial infarction prevention, 563–4
 - recurrent venous thromboembolism prevention, 562–3
- antidiuretic hormone, 46
- aplastic anemia/CCS, 5, 223
- Apotek AB, 389
- Apoteksbolaget (Sweden), 389

- apparent volume of distribution, 48
 aprindine, 186
 ARIES (Adverse Reaction Identification Evaluation System), 202
 Arthur reaction, 717
 artifactual association, 19
 asbestos, 132
 aspirin, 44
 myocardial infarction association, 212
 Reye's syndrome and, 76, 223
 Associated Permanent Workshop of Imputologists (APWI), 528
 Third International meeting, Paris (1992), 528, 535
 association, 18
 artifactual, 19
 criteria for causal nature, 20–2, 20
 types, in a study, 18–20, 19
 astemizole, 7, 56, 94
 asthma quality of life questionnaire (AQLQ), 607
 atracurium, 203
 attention deficit disorder (ADD), 87
 audit, 492–3
 autologous bone marrow transplantation (ABMT), 594–5
 Automated Inpatient Medication System (RAINS), 267
 azapropazone, 371, 647
 azithromycin, 83
 AZT, *see* zidovudine

 bambuterol, 242
 BARDI (Bayesian Adverse Reaction Diagnostic Instrument), 533–4, 536
 barium allergy, 733
 barium enema, 733–4
 Bayesian Adverse Reaction Diagnostic Instrument (BARDI), 533–4, 536
 Bayesian analysis, 718
 Bayesian causality assessment, 455
 BCG vaccine, 186, 565
 Belmont Report, 419
 Bendectin, 7, 92, 95, 100, 128, 130, 132, 133, 136, 678, 752
 litigation, 143–5
 beneficial drug effects, 98, 409, 553–72
 applicability of approaches, 560–1
 clinically important information, 554, 555
 confounding and, 558–60
 cost-effectiveness studies, 565
 examples, 561–6
 generic vs brand name drugs, 564
 post-marketing factors determining, 555–7, 556
 vaccines, 565
 Benign Breast Disease Registry: Pilot Phase, 268
 benoxaprofen, 7
 benzodiazepines, 46, 95, 758
 benzyl alcohol, 43
 beta-adrenergic blocking agents, 44
 beta2-adrenoceptors, 43
 beta2-agonists, 43
 beta-blockers, 50, 258, 557
 coronary heart disease (CHD), 255, 256
 myocardial infarction, 617, 651
 withdrawal syndrome, 44
 beta carotene, 652
 beta error, 32
 BfArM, 177
 bias, 19, 20, 319, 765–73
 in abstraction of data, 636–7
 contraindication, 543
 detection, 773
 indication, 543, 776
 information, 219, 221, 766, 771, 779–80
 language, 636
 misclassification, 770–73
 differential, 770, 773
 nondifferential, 770, 771–73
 prevalence, 767, 771
 protopathic, 453, 767, 768, 770
 publication, 635–6, 642–4
 publicity, 769
 recall, 754–5, 759, 773
 reference, 636
 referral, 767, 768–9
 sample distortion, 768
 selection, 219, 220–1, 766, 768–70, 779
 self-selection, 769–70
 in spontaneous report, 667
 vaccine safety and, 712, 725
Bichler v. Eli Lilly and Co., 142
 biflavonoids, 56
 Bilitch registry, 743
 bioavailability, 55
 biological inference, 18
 birth defects, *see* drug-induced birth defects
 Birth Defects Study, 759
 Boston Collaborative Drug Surveillance Program (BCDSP), 7, 194, 210, 376
Boyles v. American Cyanamid Co., 133, 136
 Bradford-Hill criteria, 452
 breast augmentation devices, 96
 Breast Cancer Family Registry: Pilot Phase, 268–9
 Breast Cancer Screening Program (GHC), 253
 bretylium, 65, 560
 British National Formulary, 178
 bromfenac, 7, 117
 bromocriptine, 7
 Burroughs Wellcome Foundation, 814

 caffeine, 54
 Calcipopen, 185
 calcium channel blockers, 7, 43, 222, 257–8, 318
 Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures (CCP), 329
 Canadian Institute for Health Information (CIHI), 328
 case mix group (CMG), 328
 resource intensity weight (RIW), 328
 Canadian Mortality Database, 389
 Cancer Prevention Study, 786, 787
 Cancer Surveillance System, 253
 cancer screening, efficacy, 565–6
 captopril, 604
 carbamazepine, 47, 56
 Carstairs deprivation score, 364
 case-control studies/surveillance (CCS), 24, 25–6, 209–30
 ad hoc, 392
 birth defects, 223
 confounding, 219, 221
 consistency, 219–20, 221
 data collection form, 214–15
 data sources, 391–5
 hypothesis generation and testing, 215, 222–3
 in-depth analysis, 215
 information bias, 219, 221
 methods, 213–15
 misclassification of exposure and outcome, 218–19, 220
 sample size calculations, 35–7, 36, 38–9
 selection bias, 219, 220–1
 statistical power, 216–18, 220
 strength of associations, 218, 220

- strengths, 215–20
- weaknesses, 220–1
- case–noncase analysis, 118
- case reports, 22–3
 - causation from, 525–38
- case series, 23
 - sample size calculations, 37–8
- causality assessment, terminology, 179
- causation, 18, 528–38
- ceftriaxone, 83
- Center for Devices and Radiological Health (CDRH) (FDA), 733
- Center for Drug Evaluation and Research (CDER) (FDA), 163, 733
- Center for Epidemiologic Studies–Depression scale, 268
- Center for Health Care Policy and Evaluation, 296
- Center for Health Research, 266–7, 269
- Center for Health Studies' Data Warehouse, 249
- Center for Medication Monitoring, University of Texas Medical Branch, Galveston, 393–4
- Centers for Disease Control (CDCs) (US), 87, 141, 390, 815
- Centers for Disease Control and Prevention (CDCs), 279, 713
- Centers for Education and Research I Therapeutics (CERTs), 813
- central compartment, 48
- cephalosporin, 198
- changeability, 606
- channeling, 355–7, 543, 776, 802
- chemotherapy in children, 96
- Chlamydia trachomatis* infection, 251
- chloral hydrate, 198
- chloramphenicol, 5, 153, 187
- chlordiazepam, 198
- chloromycetin, 129, 130, 153
- chlorpromazine, 200
- Chronic Respiratory Questionnaire, 606, 607
- Churg–Strauss syndrome, 114
- cigarette smoking, 22, 24, 198, 200
- cimetidine, 4, 23, 37, 56, 95, 203, 394, 395
- ciprofloxacin, 83
- cisapride, 7, 94
- clarithromycin, 301–2
- clavulanic acid, 380–1
- clear cell adenocarcinoma of the cervix and vagina, 6, 10, 22, 24, 95
- clearance (Cl), 48
- Cleveland Clinic Lead Registry, 743
- Clinic Appointment System (CAS), 269
- clinical pharmacology
 - definition, 4–5, 41
 - vs pharmacoepidemiology, 41–2
- Clinical Pharmacology and Therapeutics, 809
- clinical practice guidelines, 490–1
- clinical trials
 - clinical evaluation and patient follow-up, 79
 - disease conditions under study, 78
 - dosage, 78
 - drug interactions, 78
 - length of exposure, 77
 - number of patients, 77
 - representativeness of target population, 77, 78
- clomipramine, 54
- Clostridium difficile*, 302–3, 291, 302–3
- clozapine, 202
- Cochrane Collaboration on Effective Professional Practice, 484
- Code of Federal Regulations, 450
- cognitive theory of autobiographical memory, 662–3
- cohort studies, 24–6, 787–8
 - ad hoc*, 394–5
 - birth defects and, 757–9
 - data sources, 392–3
 - fixed-entry, 791
 - future, 802–3
 - intensive hospital-based, *see* intensive hospital-based cohort studies
 - partial confounder data, 788–9, 796–8
 - phase 4, 64
 - sample size calculations, 31–5, 34, 38–9
 - sampling within, 787–8, 790–96
 - case–cohort design, 794–6
 - nested case–control design, 792–4
 - types and structures, 790–2
 - variable-entry, 791
 - within-subject designs, 789–90, 798–802
 - case–crossover design, 799–800
 - case–time–control design, 801
 - drug database designs, 801–802
- coincidental event, 716
- Collaborative Perinatal Project (CPP) (US), 757
- collateral estoppel*, 142
- Commission on Professional Hospital Activities–Professional Activity Study (CPHA–PAS), 251
- Committee for Proprietary Medicinal Products (CPMP), 119–20
 - Pharmacovigilance Working Party, 120
- Committee of the International Organization of Medical Societies, 421, 427
- Committee on Safety of Drugs (CSD) (UK), 231
 - Serious Adverse Drug Reactions List, 218
- Committee on Safety of Medicines, 6, 231–2
- Common Control Pool (CCP), 269
- Common Recruitment Pool (PRP), 269
- Common Rule (US), 420, 421, 422, 423, 427, 428
- Common Services Agency
 - Pharmacy Practice Division (PPD), 362, 371
- Community Health Index (CHI) number, 362–3
- Community Health Master Patient Index, 362, 363, 364, 365
- Community Health Services System, 253
- Community of Tierp Project, 470
- COMPASS, 307, 309–11, 318, 471, 564
- Compassionate Use Programs, 449
- completeness, 665
- Comprehensive Hospital Drug Monitoring Berne (CHDMB), 201
- Computerized Online Medicaid Analysis and Surveillance System, 8
- Computerized On-Line Medical Pharmaceutical Analysis and Surveillance System (COMPASS), 8, 307, 309–11, 318, 471, 564
- confidence intervals, 26
- confidence profile method, 640
- confounder
 - data, 788–9
 - definition, 19
 - important, 410
 - partial, 788–9, 796–8
- confounding, 547, 557–8, 559, 773–84
 - beneficial drug defects and, 558–60
 - birth defects, 766
 - case–control surveillance and, 219, 221
 - by comedication, 777
 - control, 408
 - dealing with, 780–83
 - at analysis level, 781–83
 - at design level, 780
 - large drug databases, 782–3
 - and effect modification, 777–9
 - future, 783–4
 - by indication, 453, 543, 777
 - mechanism, 773–5
 - in random controlled trials, 547
 - by reason for prescription, 775–6

- confounding (*continued*)
 by severity, 553, 776
 vaccine safety, 712, 725
 confounding variable, 19, 19, 20
 Congenital Malformation Register (Finland), 396
 construct validity, 605
 consumer (patient) behavior, 59–60
 Consumer Project on Technology, 438
 Continuing Care Service Database, 268
 contract theory, 126
 contraindication bias, 543
 Cooperative Agreement Program in Pharmacoepidemiology (FDA-WHO), 115, 118, 119
 Coronary Drug Project, 777
 COSTART, 715
 cost–benefit analysis, 576–7
 cost–effectiveness analysis, 577–8
 cost-effectiveness studies, 565
 cost–identification analysis, 578
 co-trimoxazole, 188
 coumarin, 53
 Council for International Organizations of Medical Sciences (CIOMS), 102, 179, 183
 CIOMS I initiative, 183
 CIOMS II, 183
 County of Jämtland Project (Sweden), 469
 counterfeit drugs, 434–5
 Cox regression, 455
 Cox semiparametric regression, 592
 Creutzfeldt–Jakob disease (CJD), 95
 criterion validity, 605
 cross-sectional studies, 27
 Crystal City conference (1983), 528, 534
 Current Membership Study Database, 269
 Current Procedural Terminology (CPT) codes, 594
 cyclooxygenase (COX), 45
 COX-1 45
 COX-2 45
 cyclosporin, 52, 56, 57
 CYP2D19 polymorphism, 54
 cytochrome P450 family, 52–3, 111
 Cytology Database, 268

 Dalkon Shield, 129, 141, 223
 Danish Pacemaker Register, 743
 dapsone, 54
 DARTS initiative, 370
 data marts, 281
 database
 automated, 98–9
 quality of resources, 101–2
 quantity of resources, 101
 see also under names
Daubert v. Merrill Dow Pharmaceuticals Inc., 136, 137
Davis v. Wyeth, 128
 debrisoquine, 54
 defined daily dose (DDD), 112, 471–2
 Delaware Valley Case–Control Network, 394
 Dental Administrative and Clinical Tracking System (TEAM), 268
 Dental Visit Survey Database, 269
 Department of Health and Aged Care (DHAC) (Australian), 443
 depletion of susceptibles, 455
 depot-medroxyprogesterone, 7
 deregulation, 83
 descriptive studies, 27
 DESI (Drug Efficacy Study Implementation) process, 6

 desipramine, 54, 257
 detection bias, 773
 developing countries, 79–80
 regulatory agencies in, 85
 dexfenfluramine, 7, 47, 381, 382
 DHEW Task Force on Prescription Drugs (US), 506–7
 Diabetes Registry, 269
 diazepam, 20, 46, 95, 198
 breast cancer, 222
 metabolism, 53
 didanosine, 83
 diet medicines and heart valve disorders, 381–2
 diethylene glycol, 5, 79, 152, 434
 diethylstilbestrol, 6, 10, 22, 24, 95, 132, 209, 787
 litigation, 142–3
 differential misclassification, 663
 bias, 770, 773
 digoxin, 44, 46, 57, 198
 diltiazem, 56
 dioxin, 132, 134
 diphosphonates, 114
 diphtheria–tetanus–acellular pertussis (DTaP) vaccine, 711
 diphtheria–tetanus–pertussis–haemophilus influenza type b (DTPH) vaccine, 711
 direct medical costs, 579
 direct nonmedical costs, 579
 discriminative instruments, 605
 disease incidence data, 389–91
 disease states, effects of, 46
 distribution phase, 48
 Division of Drug Marketing Advertising and Communication (DDMAC) (FDA), 98
 DNA vaccines, 725
 dornase alpha, 371
 dose–response relationship, 21
 doxycycline, 257
 drug action, mechanisms, 43
 drug approval process, current, 8–9
 Drug Commission of the German Medical Profession, 177
 Drug Data Bank (Spain), 469
 Drug Efficacy Study Implementation (DESI) process, 6
 Drug Epidemiology Unit, 7
 Drug Experience Reports, 144
 drug-induced birth defects, 749–63
 alleged teratogenic drugs, 752
 biologic plausibility, 756
 case–control studies, 759–60
 “class action” teratogenesis, 752–3
 cohorts, 757–9
 datasets for other purposes, 757–8
 follow-up, 758–9
 large populations, 757
 confounding, 756
 drugs known to be teratogenic, 751
 drugs not known to be teratogenic, 751–52
 exposure, 753–5
 nonprescribed drugs, 754
 recall bias, 754–5
 integrated approach, 760
 integration of epidemiology/biology, 760–1
 legal climate, 761
 nature of birth defects and relation to drugs, 749–50
 nonprescription (OTC) drugs, 750–1, 754
 outcome, 755–6
 prediction prior to drug marketing, 750–1
 sample size, 753
 drug lag, 6
 drug potency, 43–4

- Drug Prescription Plan, 693
- drug safety programs, *see* safety, drug
- Drug Safety Research Trust, 8
- Drug Safety Research Unit (DRSU), 232
 - “green form”, 232, 233, 238
- Drug Surveillance Network, 203
- Drug Surveillance Research Unit, 8
- Drug Use Evaluation (DUE) Programs, 465
- Drug Utilization Research groups, 473–4
- drug utilization review (DUR), 464–5, 493, 505–23
 - clinical problems addressed, 511
 - conceptual framework, 506
 - currently available solutions, 513–19
 - framework for evaluating effectiveness, 513–14
 - future, 519
 - history, 506–8
 - hospital programs, 510–11
 - methodologic problems addressed, 511–13
 - availability of data, 512–13
 - barriers to randomized trials, 511–12
 - clustered observations, 512
 - potentially noncomparable study groups, 512
 - rare outcomes, 512
 - outpatient prospective, 510
 - outpatient retrospective, 509–10
 - in changing drug therapy and improving clinical outcomes, 517–18
 - effectiveness, 514–16
 - in improving clinical outcomes, 517
 - not identified, 516–17
 - process, 508–9
 - synonyms, 505
- drug utilization studies, 6
 - classification systems, 472–3
 - clinical problems, 465–6
 - current data sources, 469–71
 - data sources, 388–9
 - evolution, 467–8
 - future, 474–6
 - intervention strategies, 473–4
 - methodologic problems, 466–7
 - qualitative studies, 464
 - types of study, 464–5
 - units of measurement, 471–2
 - WHO definition, 463
- drug, definition, 42
- drug–drug interactions, 46–7
 - involving metabolizing enzymes, 56
 - involving renal elimination, 52
- DTP vaccine, 765–6
- Duke–UNC Health Profile, 608
- duration–response relationship, 21
- DURbase, 471
- DVTs, 581

- ecological studies, 23
- ECRI Universal Medical Device Nomenclature System, 739
- education materials, 490–1
- educationally influential physicians, 494
- effect modification, 776, 777–9
 - by dose or drug potency, 778–9
 - by duration of therapy, 455
- effectiveness, drug, 554
- efficacy, legal requirement, 84
- efficiency, drug, 554
- eflornithine, 83
- EGRET, 793
- elimination, effects of variation in, 50–2
 - elimination rate constant, 48
 - Emergency Department Database, 268
 - Emergency Psychiatric Service Logbook, 268
 - enalapril, 18
 - encainide, 110
 - enzyme induction, 56
 - enzyme inhibition, 56
 - eosinophilia–myalgia syndrome, 7
 - EPIC research unit, 377, 378
 - EpicCare, 267
 - EPICURE, 795
 - epidemiologic study designs, 22–7
 - epidemiology, definition, 5
 - epinephrine, 44
 - epipodophyllotoxins, 96
 - EPITOME, 795
 - epoprostenol (Flolan), 597
 - Erickson v. Winthrop Laboratories et al.*, 135, 136
 - erythromycin, 56
 - Essential Drug (ED) concept, 80
 - Essential Drugs Program, 441
 - Established Populations for Epidemiologic Studies of the Elderly, 473
 - estradiol, 49
 - estrogen, 209, 670, 672, 673–4
 - endometrial cancer, 222
 - osteoporotic fractures prevention, 561–2
 - ovarian cancer, 211, 213
 - Eurostep, 98
 - ethical issues, 417–31
 - currently available solutions, 426–8
 - ethics review boards, 420–1
 - future, 428–30
 - human subject, 420
 - informed consent, 423–4
 - methodologic problems addressed by, 424–6
 - minimal risk, 422–3
 - in pharmacoepidemiology research, 419–24
 - privacy and confidentiality, 421–2
 - ethynylestradiol, 56
 - Eudranet, 177
 - Eudrawatch, 177
 - EURO DURG, 348, 473, 475
 - European Agency for the Evaluation of Medicinal Products (EMA), 183, 349
 - European Drug Utilization Research Group (EURO DURG), 475
 - European Medicines Agency, 84
 - European Medicines Evaluation Agency (EMA), 177
 - European Parliament, 183
 - European Pharmaceutical Market Research Association (EPHMA), 472
 - European Privacy Directive, 304
 - European Society for Pharmacoepidemiology, 809
 - European Union/European Commission, 84, 426
 - Consumers Protection Directorate, 84
 - Directives
 - (65/65/EEC), 84, 232
 - (75/319/EEC), 232
 - (95/46/EC), 475
 - Industry directorate, 84
 - pharmacovigilance system, 177
 - EuroQol, 597
 - evaluative instruments, 605–6
 - experimental studies, 26
 - EZ measles vaccine mortality, 725

 - F test, 621
 - face validity (sensitivity), 605

- face-to-face educational outreach, 494–7
- fail-safe N, 643
- FDA Modernization Act (1997) (FDAMA) (US), 125, 126
- Federal Employers Liability Act, 129
- Federal Food and Drugs Act (1906), 152
- Federal Food, Drug, and Cosmetic Act (1938), 5, 125, 126, 130, 142, 153
- Kefauver–Harris Amendments, 6, 153, 555
- Medical Device Amendment (1976), 96
- Federal Tort Claims Act, 141
- feedback, 492–3
- felbamate, 117
- Feldene (piroxicam), 45, 66
- Fen/Phen, 256
- fenfluramine, 7, 47, 256, 381, 382
- fenoprofen, 647
- fenoterol, 7, 356–7
- Ferebee v. Chevron Chemical Co.*, 134, 136
- Fertility and Maternal Health Drugs Advisory Committee (FDA), 100
- fialuridine, 114
- finasteride (Proscar), 97, 242
- Finnish Cancer Registry, 396
- first pass metabolism, 55
- flecainide, 54
- Flolan International Randomized Survival Trial (FIRST), 597
- flosequinan (Manoplax), 110, 164
- flumazenil, 47
- flunarizine, 354
- fluoroquinolone, 116
- 5-fluorouracil, 277
- fluoxetine (Prozac), 7, 94, 257
- flurazepam, 197
- Food and Drug Administration (FDA), 5, 6, 8, 84, 91, 420, 450, 451
- Adverse Drug Reaction System, 713
- Adverse Event Reporting System (AERS), 152, 153, 155, 162–3, 168–9
- Advisory Committee, 94
- Center for Devices and Radiological Health (CDRH), 733
- Center for Drug Evaluation and Research (CDER), 163, 733
- Cooperative Agreement Program in Pharmacoepidemiology, 115, 118, 119
- Division of Drug Marketing Advertising and Communication (DDMAC), 98
- Fertility and Maternal Health Drugs Advisory Committee, 100
- Freedom of Information Office, 163
- funding, 813
- litigation, 129–30
- Medical Device Product Code, 739
- MedWatch to Manufacturer (MMP) Program, 157
- role in litigation, 125–6
- see also* Spontaneous Reporting Systems
- formula/ries, 506
- Framingham Heart Study, 589
- Fred Hutchinson Cancer Research Center, 248
- Cancer Surveillance System (CSS), 253
- Freedom of Information Act, 137, 139
- Freedom of Information and Protection of Privacy Act (Saskatchewan), 337
- Freedom of Information Office (FDA), 163
- French Medicine Agency, 177
- Frye v. United States*, 136
- funnel plot, 642–3, 643, 644
- furosemide, 46, 197
- Fyn County registry, Denmark, 743
- gene patenting, 86
- General Practitioners Research Database (GPRD), 99, 115, 375–86, 397, 408
- database, 406
- diagnosis and hospitalization data, 694
- drug file, 377
- event file, 377–8
- findings file, 378
- medical records, 378
- particular applications, 380–2
- registration file, 377
- representativeness of data, 406
- strengths, 378–9
- access to original medical records, 379
- population-based data, 378
- population large enough to study rare diseases, 378
- quality of the diagnosis information, 378–9
- validity of drug data in, 694
- weaknesses, 379–80
- additional computerized information, 379–80
- additional research units, 380
- completeness of data, 379
- linkage to other health care databases, 380
- generic vs brand name drugs, 564
- Generics Pharmaceutical Industry Association (GPIA) (GS), 84
- gentamicin, 4
- Glasgow Outcome Scale (GOS) score, 59
- Glaxo, 86
- Glaxo Wellcome, 85, 86
- glibenclamide, 188
- global burden of disease (GBD) concept, 80
- Global Harmonization Task Force, 735, 736
- global introspection approach, 718
- globalization, 79–80
- glomerular filtration, 51
- glucose-6-phosphate dehydrogenase deficiency, 4
- glutethimide, 752
- gold standard, 664
- Good Pharmacovigilance Practices (France), 178
- Group B Streptococcal (GBS) disease, 251
- group education, 492
- Group Health Cooperative of Puget Sound, 247–62, 396, 406, 470
- databases, 249–54, **250**
- Cancer Surveillance System, 253
- cause of death, 253
- Community Health Services System, 253
- enrollment, 250, **250**
- hospital, 251
- laboratory, 251
- outpatient visits, 251–3
- pharmacy, 251
- radiology, 251
- Utilization Management/Cost Management Information System (UM/CMIS), 253–4
- diagnosis and hospitalization data, 695
- disease severity, 260
- particular applications, 256–60
- strengths, 254–5
- use of database, 258–60
- completeness of database, 258–9
- indication for use, 259–60
- use of GHC data, 256–8
- validity of drug data in, 691–2
- weaknesses, 255–6
- Gruppo di Ricerca sulla Epidemiologia del Farmaco, 202
- Guillain–Barré syndrome, 117, 141, 185, 188, 708, 712
- Gulf War syndrome, 724

- Haemophilus influenzae* type b polysaccharide vaccine, 565, 710
- half-life, drug, 48
- haloperidol, 54
- halothane hepatitis, 54
- Harvard Pilgrim Health Care (HPHC), 115, 259, 285–93, 406, 408, 744
- adherence to drug therapy evaluation, 290–1
 - ambulatory encounter records, 286
 - automated records, 286–7
 - demographics of members, 286
 - diagnosis and hospitalization data, 696
 - drug effectiveness evaluation, 290
 - emergency room visits, hospitalizations records etc., 286–7
 - future, 292–3
 - identification of cohort requiring intensive review, 292
 - linkage to other data sources, 287
 - membership records, 286
 - overview, 285–6
 - particular applications, 289–92
 - physicians' behavior study, 292
 - potential adverse drug reactions evaluation, 289–90
 - strengths, 289
 - use of data for research, 287–9
 - coded diagnosis information, 287
 - diagnosis and procedure data, 287–9
 - full-text medical records, 287–9
 - pharmacy data, 287
 - validity of drug data in, 692
 - weaknesses, 289
- Harvard Canguard Medical Associates (HVMA), 285–6
- HCIA Inc., 94
- Health Action International, 436, 438
- Health Care Financing Administration (US), 811
- Health Evaluation through Logical Processing (HELP) System (Salt Lake City), 397
- Health Examination Survey, 390
- Health For All (HFA) By The Year 2000, 80
- Health Information Designs, Inc. (HID), 309, 471
- Health Information Protection Act (1999) (Saskatchewan), 341
- Health Insurance Portability and Accountability Act (HIPPA), 304
- health maintenance organizations (HMOs), 98, 247, 296, 514, 811
- Health Plan Employer Data and Information Set (HEDIS), 296, 700
- health profiles, 608–9
- Health Records Survey (US), 390
- health related quality of life (HRQL)
- clinical problems addressed, 604
 - definition, 603–4
 - generic instruments, 608–9
 - health profiles, 608–9
 - in investigating new drugs, 605–8
 - methodologic problems addressed, 604–5
 - multiple measures, 610–11
 - specific instruments, 609–10
 - taxonomy and potential use, 608
 - utility measurement, 609
- Health Resources and Services Administration Bureau of Health Professions
- Area Resource File (ARF), 743
- health sector reform, 83
- Health Utilities Index, 609
- Healthcare Cost and Utilization Project (HCUP), 743
- Nationwide Inpatient Sample (NIS), 743
- Helsinki Declaration (1964), 420
- heparin
- intraventricular hemorrhage, 541
 - meta-analysis, 639
- hepatitis B vaccine, 707
- herbal medications, 812
- herd immunity, 709
- hexachlorophene toxicity, 49
- Hib vaccines, 721
- Highlights of Drug Utilization in Office Practice, 389
- histamine-2 receptor antagonists (H2Ras), 514
- Histology Database, 268
- HIV ADR reporting scheme (UK), 189
- Hodgkin's disease, 96, 134, 199
- homologation, drug, 85
- Hospital Discharge Register (Finland), 396
- Hospital Discharge Survey (US), 390
- Household Interview Survey Database, 269
- hydralazine, 54, 188, 222
- hydromorphone, 43
- hyperprescribing, 357–8
- hypersensitivity reactions, 4
- hypoglycemia, 7
- hypothesis generating studies, 67, 199, 209, 211, 408–9
- hypothesis strengthening studies, 409
- hypothesis testing, 66–7, 198–9, 209, 211, 409
- ibuprofen, 36, 539, 542–3
- meta-analysis, 647
 - Reye's syndrome, 542
- ICD system, 694–5
- ICD-9-CM coding scheme, 315–16
- ICI-Pharma, 202
- IL-3/GM-CSF, 594–5
- imipramine, 257
- Immunization Database, 268
- Immunization Monitoring Program–Active (IMPACT), 715
- immunoglobulin, intravenous, and aseptic meningitis syndrome, 169–70
- immunologic reactions, 4
- Implantable Lead Registry, 743
- IMS America, Ltd, 470
- MediPlus database, 397
 - National Prescription Audit, 469
- IMS Health, 94, 112, 388
- In re: "Agent Orange"*, 133, 134
- incidence density, 242
- for meloxicam, 243
- incidence of disease, 409–10
- IND, 450–1
- indication bias, 543, 776
- Indicator Measurement System, 811
- indinavir, 83
- individual practice associations (IPAs), 247
- indomethacin, 7, 355
- information bias, 219, 221, 766, 771, 779–80
- information, patient, 120
- Inpatient Admission/Discharge/Transfer System (ADT), 267
- Institute of Medicine (US), 707
- Institutional Population Survey (US), 390
- institutional review board (IRB), 417
- Instrumental Activities of Daily Living, 268
- insulin, human, 7
- intangible costs, 580
- Integrated Primary Care Information (IPCI) system, 350
- Integrated Services Digital Network (ISDN), 371
- intensive hospital-based cohort studies, 193–207
- applications, 196–202
 - description, 194
 - descriptive studies of adverse drug reactions, 197
 - drug utilization studies, 196–7
 - hypothesis generating studies, 199

- intensive hospital-based cohort studies (*continued*)
 hypothesis-testing studies, 198–9
 other studies, 200–2
 prehospital drug use of inpatients, 199–202
 strengths, 194–5
 weaknesses, 195–6
- Intercontinental Medical Statistics International (IMS), 186
- interferon alfa-2a, 42
- International Agranulocytosis and Aplastic Anemia Study (IAAAS), 320, 771
- International Classifications of Disease (ICD)
 ICD-8, 391
 ICD-9, 391
- International Clinical Epidemiology Network (INCLEN), 475, 813
- International Codes of Ethics, 423
- International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), 84, 113, 183, 450
- International Conference on National Medicinal Drug Policies (1995), 441
- International Conference on Pharmacoepidemiology, 809
- International Federation of Pharmaceutical Manufacturers Associations (IFPMA), 84
- International Monetary Fund, 83
- International Network for Rational Use of Drugs (INRUD), 471, 475
- International Non-proprietary Names (INN), 84
- International Pharmaceutical Market Research Group (IPMRG), 472
- International Society for Pharmacoepidemiology (ISPE), 8, 102, 427, 428, 475, 809
- International Society of Epidemiology (ISPE), 304
- interpretability, instrument, 606
- interrupted time-series designs, 486
- intravenous immunoglobulin (IVIG), 169–70
- Investigational New Drug Application (IND), 6, 450, 529
- Iowa Drug Information System (IDIS), 473
- ipratropium bromide, 618, 621–2
- IRB, 421
- isoniazid, 53–4, 198
- isoproterenol, 45
- isotretinoin (Accutane), 7, 164, 751, 753
- Italian National Health Service, 397
- Japanese encephalitis vaccine, 565, 718, 721
- Johns Hopkins Hospital and Boston Collaborative Drug Surveillance Program, 6
- Johnston v. The United States*, 132
- Joint Commission for Accreditation of Hospitals, 203
- Joint Commission on Accreditation of Healthcare Organizations (JCAHO), 157, 465, 508, 811
- Joint Commission on Prescription Drug Use, 7–8, 248
- Journal of Clinical Epidemiology*, 809
- Journal of Pharmacoepidemiology*, 811
- Kaiser Appointment/Registration/Encounter System (KARE), 267
- Kaiser Permanente Interregional Genetics Registry, 268
- Kaiser Permanente Medical Care Program, 263–83, 470
 California Region, 264–6, 562
 case-control studies, 277
 clinical organization, 264
 Division of Research, 264–5
 initial phase, 272
 pharmacoepidemiology in, 265–6
 size of database, 406
 surveillance for carcinogenic effects (1969–73), 272–6
- diagnosis and hospitalization data, 695–6
 future, 280–1
 Northwest Division (KPNW), 248, 266–70
 applications, 277–80
 biologics and devices, 279
 Center for Health Research Databases, 269
 clinical and dental data systems, 267–8
 clinical organization, 266
 databases, 267–70
 disease registries, 268–9
 drug use in pregnancy, 278–9
 methodologic work, 279–80
 pharmacy database, 270, 270
 psychotropic drugs, 278
 regulatory questions, 277–8
 size of database, 406
 particular applications, 271–80
 population-based data, 407
 representativeness of data, 406
 strength, 270–1
 validity of drug data in, 692
 weaknesses, 271
- Kävepenin, 185
- Kefauver–Harris Amendments (1962) (US), 6, 153, 555
- ketoprofen, 356, 647
- ketorolac, 7, 66, 394
- Kolmogorov–Smirnov test, 591
- Kracker v. Spartan Chemical Co., Inc., et al.*, 136
- Kramer v. Wilkins*, 137
- Laboratory Information System (LIS), 268
- lamivudine, 83
- language bias, 636
- LAREB, 352
- large simple trials (LSTs), 545–6
- latex allergy, 734
- learned intermediary rule defense, 128
- least squares regression, 591
- leprosy vaccine, 565
- leucotriene receptor antagonists, 114
- levodopa, 46
- lidocaine, 563, 648
- LifeLink, 388
- lithium, 47, 57
- litigation, US
 causality assessment, 130–2
 compliance with statute to product liability litigation, 129–30
 drug product liability defense, 128–9
 expert testimony, 132–7
 FDA role in, 125–6, 129–30
 impact on pharmacoepidemiology, 140–1
 plaintiff's burden of proof, 126–8
 regulation versus, 125–30
 spontaneous reports, 137–8
 use of ADE reports, 138–40
- losartan, 371
- low incidence outcome, 410
- low prevalence outcome, 410
- L-tryptophan, 7, 170–1
- MacDonald*, 128
- managed care organizations (MCOs), 247
- Managing Drug Supply*, 434
- Manitoba case-control data, 391–2
- Manoplax (flosequinan), 110, 164
- Mantel–Haenszel odds ratio, 782
- Mantel–Haenszel procedure, 639–40, 799
- Marketing Authorization Holder (MAH), 177, 183, 185

- Massengill Company, 5
 matching, 780
 maximum likelihood estimate, 641
 Mayo Clinic, 391, 425, 744
 Rochester Epidemiology Project, 425
 McGill Hodgkin's Disease, 787
 McMaster Health Index Questionnaire, 608
 measles vaccine, 565
 measles-mumps-rubella vaccine, 255, 257, 707, 719
 measurement error, 663-4
 quantitative indices, 664-7
 MEDDRA, 163, 183
Médecins Sans Frontières, 83, 438
 Medicaid, 8, 98, 249, 296, 406, 408, 425
 appropriate role of data in pharmacoepidemiology, 321
 databases, 307-24
 diagnosis and hospitalization data, 696-7
 examples of studies, 320-1
 future, 321-2
 strengths, 314-15
 validity of drug data in, 693
 weaknesses, 315-20
 confounding variables, 316-17
 data validity, 317-19
 diagnostic terminology, 315-16
 eligibility changes, 319-20
 generalizability, 315
 potential biases, 319
 Medicaid Management Information System (MMIS), 308, 469
 Medical Device Amendment (1976) (US), 738
 Medical Device Product Code (FDA), 739
 medical devices, 733-47
 breast implants, 7, 96-7
 combinations of techniques, 744
 definition, 734-7
 differential effects, 738
 epidemiologic resources, 745
 epidemiological trends, 745
 individual exposure assessment, 738-9
 individual safety, 737-8
 medical care claims, 743
 medical records, 743-4
 national population exposure assessment, 739-40
 new data collection, 744
 public health impact, 738
 recognition of new syndromes/adverse events, 738
 registries, 743
 safety surveillance, 740
 surveys, 740-3
 technological trends, 745
 Medical Dictionary of Regulatory Activities (MedDRA), 163, 183
 Medical Outcomes Study, 608
 Medical Therapeutics and Outcomes Project (MTO), 296
 Medical Visit Survey Database, 269
 Medicare, 249, 256, 296, 392, 407, 811, 813
 Medicare Plus II Database, 268
 Medicare Prospective Payment System, 583
 Medicines Act (1968), 231
 Medicines Control Agency (MCA), 99, 115, 118, 183
 Medicines Monitoring Unit (MEMO), Tayside, 361-74
 accessibility to medical records, 366
 CHI numbers, 362-3, 371
 diagnosis and hospitalization data, 697
 drug data, 363, 365-6
 drug safety research, 369-70
 drug utilization research, 367-9
 future, 371-2
 hospital data, 363-4
 methodological studies, 370
 other in-hospital data, 364
 particular applications, 367-71
 patient identification, 362, 365
 population based data, 365
 prescriber data, 363
 prescription drug data, 362
 primary care data, 364
 strengths, 365-6
 validation of hospital data, 364
 validity of drug data in, 694
 weaknesses, 366-7
 MediPlus database, 397
 MEDLINE, 637
 MedWatch Form 3500s, 138, 139
 MedWatch program (FDA), 8
 MedWatch to Manufacturer (MMP) Program (FDA), 157
 MedWatch, data collection, 156-62
 meloxicam, 45
 events reported, 240, 241
 incidence density, 243
 reasons for stopping, 244
 Membership Information Processing System (MIPS), 267
 meningococcus vaccine, 565
 Menstrual and Reproductive Health Study, 688
 MER-29 cases, 129
 Merck Laboratories, 97
 Merrel, William S., Labs, 143
 meta-analysis, 633-59
 adverse effects investigation, 644-8
 benefits, 650-2
 combinability of studies, 635
 data collection, 638-9
 definition, 633
 differential effects, 649
 future, 652-4
 heterogeneity, 641-2
 inclusion/exclusion criteria, 637-8
 indications for existing therapies, 648-9
 literature search, 637
 publication bias, 635-6, 642-4
 quality of original studies, 635
 reasons for conducting, 634-5
 selection of, 649-50
 statistical analysis, 639-41
 steps involved, 637-41
 metabolism, drug, effects of variation in, 52-6
 metadata, 281
 metformin, 119, 337
 methadone, 23
 methyl dopa, 197, 604
 methylphenidate (Ritalin), 87
 metiamide, 23
 metoprolol, 54
 metronidazole, 95
 mianserin, 54
 mibefradil (Posicor), 7, 56, 93, 94
 misclassification bias, 770-3
 differential, 770, 773
 nondifferential, 770, 771-3
 moclobemide, 242
 modeling, 782
 monitored adverse reaction (MAR), 164
 Monitoring System for Adverse Events Following Immunizations (MSAEFI), 713
 monoamine oxidase inhibitors (MAOIs), 47
 Monte Carlo simulation, 802

- Morges Conference (1981), 528, 530
- morphine, 43
 presystemic clearance, 55
- morphine-6-glucuronide, 55
- Mortality Follow-back Survey, 740, 742
- multimedia warning campaigns, 491
- Multiple Project Assurance, 420
- Multivariate analysis, 782
- MVS mainframe, 249
- myalgia syndrome, 170–1
- Na^+/K^+ ATPase, 44
- N*-acetylbenzoquinoneimine, 55
- naloxone, 23, 47
- National Agency for Medicines and Social Insurance (Finland), 469
- National Ambulatory Medical Care Survey (US), 98, 389, 390
- National Birth Defects Prevention Study, 759
- National Board of Health and Welfare (Sweden), 470
- National Cancer Institute
 Surveillance, Epidemiology, and End Results (SEER) program, 253, 273–4
- National Center for Health Statistics, 97, 389, 390
- National Childhood Encephalopathy Study (NCES) (US), 711, 765–6
- National Childhood Vaccine Injury Act (1986) (US), 713
- National Committee for Quality Assurance, 280
- National Corporation of Pharmacies (Apoteket) (Sweden), 389, 469, 470
- National Death Index, 389
- National Disease and Therapeutic Index (NDTI), 112–13, 388, 389, 470
- National Drug Code (NDC), 286, 309, 473, 564
 Health Information Services, 94
- National Health and Nutrition Examination Survey, 740
- National Health Interview Survey (1988)
 Medical Device Implant Supplement, 740
- National Health Service, 86, 87, 113
 Information and Services Division (ISD), 371
- National Home and Hospice Care Survey, 740
- National Immunization Days, 725
- National Institute of General Medical Sciences, 813, 814
- National Institute of Health (Bundesgesundheitsamt), 177
- National Institutes of Health (NIH), 97, 126, 169, 420, 813–14, 815
- National Maternal and Infant Health Survey, 740
- National Medicinal Drug Policies, 433–47
 clinical problems addressed, 434–8
 defining, 438–9
 economic strategies, 442–3
 future, 444
 implementation, 440–3
 lack of availability of essential drugs, 435–6
 legislative and regulatory frameworks, 441
 objectives, 439–40
 poor quality and counterfeit drugs, 434–5
 poor quality use of pharmaceuticals, 436–7
 program for achieving quality use of medicines, 442
 systems for choosing, 441–2
 unintended effects of international activities, 437–8
- National Natality and Mortality Surveys (US), 390
- National Poison Centers, 113
- National Prescription Audit, 469
- National Prescription Audit Plus (NPA), 112, 113, 388
- National Technical Information Service, 163
- National Transportation Safety Board, 8
- National Vaccination Registry, 723
- nedocromil, 242
- negligence *per se* theory, 129, 130
- Netherlands Pharmacy Record Linkage, 347–59, 406
 channeling, 355–7
 community pharmacy system, 351
 diagnosis and hospitalization data, 697
 dimension of drug exposure and its correlates, 354–5
 dimension of time in drug exposure, 353–4
 future, 358
 general practice system, 350
 hyperprescribing, 357–8
 medical record linkage, 351–2
 strengths, 352
 system of medicine, 348–50
 weaknesses, 352–3
 validity of drug data, 693–4
- network model HMOs, 247
- neural tube defects, 755
- nevirapine, 83
- New Drug Applications (NDAs), 8, 92, 94, 139, 153, 450
- New Jersey Medicaid database, 307, 314
- New York Heart Association, 598
- nicotine, 49
- nifedipine, 49, 258
- nitrazepam, 54
- nitroglycerin, 49
- nitroprusside, 560
- nitrosamines, 273
- NIVEL study, 350
- N*-of-1 randomized clinical trials, 615–31
 autocorrelation of data, 622–3
 clinical interpretation, 624–5
 consent form, 630
 design, 616–17, 617
 future, 629–30
 nonparametric tests, 519–20
 outcome measures, 617–18
 parametric tests, 620–2
 practical aspects, 625–6
 role of, 627–9
 assessing potential drug impact, 628–9
 measurement of outcome, 628
 open run-in period, 627–8
 optimizing dose, 628
 predicting response, 629
 rapidity of onset and termination of action, 628
 statistical interpretation, 618–23
 visual inspection of data, 618–19
- nomograms, 31
- noncompliance, 812
- nondifferential misclassification, 663, 664
 bias, 770, 771–3
- nonexperimental study designs, 27
- nonparametric hazards models, 592
- nonparametric tests, 519–20
- norepinephrine, 44, 46
- nortriptyline, 54
- Norwegian Drug Monopoly, 469
- Norwegian Medicinal Depot, 472
- Nottingham Health Profile (NHP), 597, 608
- Novak v. United States*, 133
- NSAIDs, 42, 44, 45, 46, 52, 320, 338, 669
 bowel cancer and, 222
 choice of study approach, 412–13
 meta-analysis, 634, 645–8
- Nurses' Health Study, 786, 787
- observational study designs, 27
- oculomucocutaneous syndrome, 6, 115, 185

- odds ratio, 25, 641
- Odense Pharmacoepidemiologic Database (OPED), 469
- Office of Extramural Research, 428
- Office of National Statistics (UK), 97, 376
- Office of Postmarketing Drug Risk Assessment (OPDRA), 162, 163, 164, 165, 172
- Ohio Medicaid database, 311
- Olmstead County Medical Records, MN, 391
- Olmstead County Study of Urinary Symptoms and Health Status Among Men, 97
- omeprazole, 56
- Omnibus Budget Reconciliation Act (1990) (US), 465, 508
- Omniflox (temafloxacin), 7, 94, 113–14, 170
- one-group, post-only design, 486
- one-group pre-post design, 487
- one-step method, 639
- opiates, 46
- opinion leaders, 494
- opportunity cost, 576
- Options Health Care, Inc., 248
- Oralflex cases, 129
- oral contraceptives, 7, 23, 209, 210, 211
 - breast cancer, 648, 663–4
 - choriocarcinoma, 222, 224
 - malignant melanoma, 222
 - rheumatoid arthritis, 347–8
 - validity of data, 670, 671
- Organisation for Economic Co-operation and Development, 418
- ORYX system, 811
- Oscorel, 356
- Osmosin, 355
- osteoporosis/esophageal events, 301
- Outpatient Pharmacy System (TOPS), 267
- Outpatient Summary Clinical Record (Kaiser), 695–6
- Outpatient Utilization System (OPUS), 269
- Outside Claims and Referral System (OS-CAR), 267
- over-the-counter drugs
 - adverse experience reporting requirements, 154, 156
 - birth defects and, 750–1, 754
- Oxendine*, 145
- Oxford Community Health Project, 396
- Oxford Family Planning Association, 787
- Oxford Record Linkage Study, 396
- OXMIS coding system, 377
- PAID Prescriptions, 507
- paired *t*-test, 621–2
- panel reactive antibody (PRA), 649
- Paoli Railroad Yard PCB Litigation*, 134, 136
- parametric failure-time models, 592
- Parke-Davis, 130
- paroxetine, 54
- passive diffusion, 49
- passive tubular reabsorption, 51
- Pasteur Merieux, 435
- penicillin, 22, 57, 201
- pentoxifylline, 256
- Periodic Safety Update Report (PSUR), 121–2
- peripheral compartment, 48
- perphenazine, 54
- pertussis vaccine, 565
- Peteet v. The Dow Chemical Co.*, 134
- Pfizer Pharmaceuticals, 66
- Pharmaceutical Benefits Advisory Committee (PBAC) (Australia), 443
- Pharmaceutical Benefits Scheme (PBS) (Australia), 442, 443
- pharmaceutical industry
 - epidemiology support
 - in marketing, 97–8
 - in new product planning/portfolio development, 97
 - evaluating drug safety, 92–6
 - goals and objectives, 91–2
 - mergers, 85–6
- Pharmaceutical Research and Manufacturing Association (PhRMA), 91, 102
- pharmacodynamics
 - definition, 4, 44, 76
 - in determining variability of drug response, 44
 - effect of age on, 45–6
 - effects of disease states, 46
- pharmacoeconomics, 573–601
 - currently available solutions, 593–8
 - economic evaluation and drug development process, 575–6
 - future, 598–9
 - measurement and modeling, 589–93
 - analysis of costs methods, 591–3
 - analysis plan for cost data, 590–1
 - multivariable analysis, 592–3
 - uncertainty in economic assessment, 593
 - univariate analysis, 591–2
 - methodology in assessment of therapies, 580–1
 - perspective of analysis, 580
 - problems, 581–3
 - research trends, 574–5
 - resource consumption, 588–9
 - study design, 583–8
 - appropriate comparators, 585
 - data collection, 585
 - multicenter evaluations, 585–8
 - participation of patients, 584–5
 - sample size, 584
 - techniques of clinical economics, 576–80
 - types of analysis, 576–9
 - types of costs, 579–80
- Pharmacoepidemiologic Prescription Database of the County of North Jutland, 469
- Pharmacoepidemiology and Drug Safety*, 809
- pharmacoepidemiology data resources
 - comparative characteristics, 406–12
 - control of confounding, 408
 - effect of research questions on, 408–12
 - inpatient drug use data, 408
 - outpatient diagnoses data, 408
 - population-based, 407
 - relative cost, 406
 - relative size of database, 406
 - relative speed, 406
 - representativeness, 406–7
 - validity of exposure data, 407
 - validity of outcome data, 407–8
- pharmacoepidemiology studies
 - aims of, 75–9
 - choosing approach to, 401–6
 - definition, 3–5
 - education and, 102
 - general contributions, 11
 - historical background, 5–8
 - logistics and utility, 100–1
 - misuse or misinterpretation of studies, 99–100
 - Phase I, 72, 76
 - Phase II, 72, 76
 - Phase III, 76–7
 - potential contributions, 9–11, 9
 - premarketing studies, 10–11

- pharmacoepidemiology studies (*continued*)
 reasons to perform, 63–7, **64**
 supplementary information, 9–10
 vs clinical pharmacology, 4–5
 vs epidemiology, 5
- pharmacokinetic compartment, 48
- pharmacokinetics
 consequences of variability, 57
 definition, 4, 44, 76
 in determining variability of drug response, 47
 effect of age on, 45
 mathematical parameters used in, 47–9
- pharmacotherapeutics, definition, 41, 42
- Pharmaco-Therapy Discussion groups, 358
- Pharmacy Assistance Aging Program, 314
- Pharmacy Information Management System (PIMS), 265, 276–7
- Pharmacy Practice Division (PPD) (CSA), 362, 371
- PHARMO-system, 347, 351–2, 407, 697
- phenformin, 119
- phenobarbital, 198
- phenothiazines, 46, 273
- phentermine, 47, 256, 381
- phenylbutazone, 7, 37, 53
- phenylpropanolamine, 540
- phenytoin, 56, 198, 751
- pHisoHex, 145–6
- phocomelia, 6
- Physicians' Health Study, 548
- picloram, 133
- piroxicam (Feldene), 45, 66
- placebo controls, 10
- pneumococcal vaccine, 565
- point of service (POSS) arrangements, 296
- Poison Information Centers, 177
- Poisson regression, 455
- poliovirus vaccine, oral, 565, 708
- polychlorinated biphenyls (PCBs), 132, 134
- population based care, 253
- Porter v. Whitehall Laboratories, Inc. et al.*, 136
- Posicor (mibefradil), 7, 56, 93, 94
- postmarketing surveillance studies, 5, 232
 data sources, 393–4
- postmarketing surveillance system, 83
- potassium chloride, 197
- practolol, 6, 115, 209
- prazosin, 10, 23, 37, 66, 394
 presystemic clearance, 55
- PRECEDE, 490
- predictive values, 665
- preferred provider organizations (PPO), 296
- pregnancy
 medication in, 278–9, 674–9
 prescription-event monitoring, 243–4
 registers, 95
 Saskatchewan studies, 339
- Pregnancy Registry, 269
- Premarket Approval Application (PMA) process, 96
- premarketing applications, 110, 449–62
 analogy, 454–5
 assessment of causality of adverse events, 452–7
 biological plausibility, 454
 coherence, 454
 consistency, 452
 currently available solutions, 457–60
 dose–response, 453–4
 experimental evidence, 454
 future, 460
 information technology requirements, 456–7
 specificity, 455
 statistical analysis methods, 455–6
 strength of association, 452
 temporality, 452–3
- pre–post with comparison group design, 486
- prescribed daily dose (PDD), 472
- Prescriber Drug User Fee Act (PDUFA) (1992), 95
- prescriber practice, 57–60, 86–7
 clinical problems, 484–5
 economic influences, 58–9
 ethical and legal problems in hindering RCTs, 488–9
 expectation and demand, 58
 financial incentives and penalties, 497
 future, 497
 influences on, 57–60
 internal validity, 485–7
 logistical issues, 488
 patient behavior and, 59–60
 patient outcomes, 57–8, 489
 pecuniary interests in prescribing, 59
 pharmaceutical industry influences on, 59
 regression toward the mean, 487–8
 reminder and decision support systems for, 493–4
 risks and benefits of treatment, 58
 unit of analysis, 488
- Prescription Drug Plan, Saskatchewan, 325, 326–7
- Prescription Drug User Fee Act (1992) (US), 8
- prescription-event monitoring, 8, 231–46
 background effects, 244–5
 description, 232–8
 display of events during the study, 239–41, **240**
 display of events on drug, 239–41, **241**
 event rates of adjusted rates, 242
 incidence density, 242
 long latency adverse reactions, 242–3
 outcomes of pregnancy, 243–4
 reasons for withdrawal, 242
 searching for signals, 239–44
 strengths, 238
 study of diseases, 245
 weaknesses, 239
- Prescription Pricing Authority (PPA), 113, 232, 377, 396, 469
- Prescription Pricing Bureau, 395
- presystemic clearance, 55–6
- prevalence bias, 767, 771
- price fixing, 85
- probenecid, 52
- procainamide, 54
- productivity costs, 579–80
- Professional Standards Review Organizations (PSRO) (US), 465, 764
- profiling, 492–3
- programmatic error, 716
- propafenone, 54
- Propensity Scores, 559
- proportional reporting ratios (PRR), 118, 180–1
- propranolol, 44, 45, 604
 presystemic clearance, 55
- Proscar (finasteride), 97, 242
- prospective study, 27
- prosthetic joint replacements, 96
- Protocare Sciences, 309
- Protocare Sciences Proprietary Medicaid Database, 309
- protopathic bias, 453, 767, 768, 770
- Provider Perspective, 113, 388
- Prozac (fluoxetine), 7, 94, 257
- public health, 79–86
- publication bias, 635–6, 642–4

- publicity bias, 769
 Pure Food and Drug Act (1906) (US), 5
p-values, 26, 620

 quality adjusted life years (QALYs), 609
 quality adjusted months (QALMs), 597
 quality of life measurements, 603–11
 Quality of Well-Being Scale, 609
 Quality Screening Management (QSM), 296
 quinidine, 648

 Radiology Information Management (RIM) Database, 268
 random effects model, 640
 random error, 19, 765
 randomization, 780
 randomized controlled trials (RCTs), 26–7, 76–7, 355, 539–52
 classic, 543–5
 blinding, 544
 choice of control treatment, 544
 data analysis, 544–5
 data collection, 544
 generalizability of results, 545
 sample size, 544
 data monitoring/interim analyses, 549
 future, 549–50
 large simple trials (LSTs), 545–6
 appropriateness of, 546–7
 data collection, 546
 data elements, 546
 feasibility conditions, 547–8, 547
 follow-up, 548
 logistics, 548–9
 power and confounding, 547
 power/sample size, 545–6
 treatment/outcome interaction, 547
 N-of-1, *see N*-of-1 randomized clinical trials
 as postmarketing surveillance studies, 395–6
 primary analysis, 548–9
 subgroup analyses, 549
 rauwolfia alkaloids, 211, 212, 222
 recall bias, 754–5, 759, 773
 reference bias, 636
 referral bias, 767, 768–9
 Regenstrief Medical Record System, 396
 Register of Persons Entitled to Free Drugs (Finland), 396
 regression to the mean, 487–8, 558
 regulation, drug, 83–4
 regulatory agencies, 109–22
 Relenza, 86
 reliability, 605
 reliability measures, 664, 666–7
 reproducibility, 606
 research synthesis, 633
 responsiveness, 605
 restructuring of design, 778
 Retail Perspective, 388
 retrospective study, 27
 Reye's syndrome, 76, 132, 223, 492, 493, 542, 766
Reyes v. Wyeth, 128
 Rheumatology Registry, 269
 Rhode Island Drug Use Reporting System, 812
Richardson, 145
Richardson v. Richardson-Merrell, Inc., 143
 rifampicin, 56
 risk
 absolute, 547
 categories of, 151, 152
 relative, 547
 vs safety, 67
 tolerance, 68–70
 characteristics of the exposure, 69
 feature of the adverse outcome, 68–9
 perceptions of the evaluator, 70
 risperidone, 339
 Ritalin (methylphenidate), 86
 Robins, A.H., Co. Inc, 141
 Rochester Epidemiology Project (Mayo Clinic), 425
 Rockefeller Foundation, 814
 rotavirus vaccine, 7
 Royal College of Physicians, 421, 427
Rubanick v. Witco Chemical Corp., et al., 134–5, 136
 rubella, maternal, 749
 rubella vaccine, 565, 717
 rule of threes, 38, 93

 Safety Appliance Act, 129
 safety, drug, 67
 adverse reaction case reporting, 113–15
 aggregate analysis of case reports, 117
 analytic uses of case reports, 117–18
 case reports and case series, 116
 clinical trials, 93
 drug usage data, 112–13
 elements of, 112–19
 formal epidemiology studies, 94–6
 future directions, 120–2
 medical devices, 96–7
 need for, 110–12
 pharmaceutical industry and, 92–6
 population based data resources, 115–16
 public health and, 79–86
 regulation, US, history of, 152–3
 signaling strategies, 118–20
 spontaneous reporting systems, 93–4
 use of population based data resources, 118–19
 use of postmarketing safety data in regulatory decision making, 119–20
 see also vaccine safety
 salbutamol, 356, 607
 Salk vaccine trial, 545
 salmeterol, 242, 607
 sample size, 31–9
 case-control studies, 35–7, **36**, 38–9
 case series, 37–8
 cohort studies, 31–5, **34**, 38–9
 tables, 817–50
 SAS, 456
 Saskatchewan Cancer Agency, 331
 Saskatchewan Cancer Program, 331
 Saskatchewan Drug Database, 99
 Saskatchewan Drug Plan, 340–1
 Saskatchewan health databases, 325–45
 alcohol and drug abuse treatment data, 334
 cancer services data, 331, **333**
 cross-sectional and longitudinal studies, 335
 data verification/validation, 335–6
 diagnosis and hospitalization data, 697
 electronically linkable data, 335
 eligible population, 325–6
 future, 341
 Health Services Number (HSN), 325
 hospital services data, 327–9
 ICD-10 coding, 335
 laboratory records, 334
 medical records, 334, 335
 mental health services, 334

- Saskatchewan health databases (*continued*)
- outpatient prescription drug data, 335
 - particular applications, 337–41
 - physicians services data, 329–31, **332**
 - population based data, 335
 - prescription drug data, 326–7
 - size of database, 406
 - strengths, 335–6
 - subject identification, 335
 - supportive care services, 334
 - validity of data, 691, 693
 - vital statistics, 331–4
 - weaknesses, 336
- Saskatchewan Health Information Network (SHIN), 341
- Saskatchewan Health Plan, 669
- Saskatchewan Health Services Utilization and Research Commission (HSURC), 340
- Scientific and Ethical Advisory Board (SEAG), 376
- scientific inference, 18
- scientific study method, 17–18, 18
- scopolamine, 49
- Scottish Morbidity Record 1 (SMR 1), 363–4
- Selacryn cases, 129
- selection bias, 219, 220–1, 766, 768–70, 779
- selective serotonin reuptake inhibitors (SSRIs), 47
- sensibility, 605
- sensitivity analysis, 578–9, 668
- sensitivity measures, 665
- Serious Adverse Drug Reactions List (CSD), 218
- serotonergic syndrome, 46
- serotonin, 47
- SF-36, 608
- Sickness Impact Profile (SIP), 607, 608
- sildenafil (Viagra), 7, 47, 256
- silicone breast implants, 7, 96–7
- Slone Epidemiology Unit, 406
- SmithKline Beecham, 85
- smoking, 22, 24, 198, 200
- SNIP criteria, 118
- Social Health Maintenance Organization (Social/HMO) Database, 269
- Social Security Administration, 580
- Social Security death tapes (US), 389
- sorivudine, 277
- sparteine, 54
- specificity, 21
- specificity measures, 665
- spermicides, teratogenic effects, 95, 100, 145
- spillover effect, 506
- spironolactone, 197
- spontaneous reporting, 137–8
- outside the US, 175–85
 - applications, 185–9
 - feedback to reporters, 180
 - generation and evaluation of signals, 180–2
 - handling and evaluation of reports, 178–80
 - international reporting, 183
 - organization, affiliation, and task of national monitoring centres, 177–8
 - rate of reporting, 182–3, 182
 - reporting requirements, 178
 - sources of reports, 178
 - strengths, 184
 - weaknesses, 184–5
 - types, 176–7
 - in the US, 151–74
 - adverse event recognition, 167
 - applications, 168–71
 - biases, 167
 - clinician contributions, 167
 - estimation of population exposure, 168
 - FDA evaluation, 163–5
 - hypothesis generation, 166
 - large-scale and cost-effective, 165–6
 - report quality, 168
 - reporting requirements, 153–6, 154
 - serious AE, 154
 - signals of problems, 166
 - strengths, 165–7
 - unexpected AE, 154
 - underreporting, 167
 - weaknesses, 167–8
 - see also* Adverse Event Monitoring System; Medwatch
 - vaccines and, 712–20
- Spontaneous Reporting System (SRS), 137–8, 139, 140, 153, 296
- standardization, 781
- standardized incidence rate (SIR), 794
- standardized morbidity/mortality ratios (SMRs), 456
- standardized mortality rate (SMR), 794
- Stanton v. Astra*, 129
- Stata, 456
- state of the art defense, 128
- statistical inference, 18
- steady state, 49
- Sterling v. Velsicol Chemical Co.*, 133
- Stevens–Johnson syndrome, 184, 186, 224, 317, 318
- Stahl* litigation, 145
- stratification, 781–2
- streptokinase, 649
- structural adjustment, 83
- Students *t*-test, 591, 621
- study design, 21
- study subjects, 17
- subacute myelo-optic-neuropathy (SMON), 6
- subarachnoid hemorrhage, 592–3
- sudden infant death syndrome (SIDS), 279, 783
- sulfalazine, 54
- sulfanilamide, 5, 152–3
- sulfasalazine, 188
- Summary of Product Characteristics (SPC), 122
- Supplemental Security Income Program (SSIP), 307–8
- sufopren, 7, 188, 391
- Surgeon General's Report on Smoking and Health (1964), 20
- Surveillance Epidemiology and End Results (SEER) program, 253, 273–4
- Swedish Diagnosis and Therapy Survey, 470
- Swedish Medical Association, 470
- Swedish Pharmaceutical Data Ltd (LSAB), 470
- Swedish Prescription Survey, 469
- swine influenza vaccine, 141, 708, 709, 718
- Sybase Data Warehouse, 249
- System for Thalidomide Education and Prescribing Safety (STEPS), 751
- systematic error, 765
- systematic review, 633
- systemic distribution, effects of variation in, 50
- Talc Morhange, 145
- temafloxacin, 116
- temafloxacin (Omniflox), 7, 94, 113–14, 170
- temafloxacin syndrome, 170
- temazepam metabolism, 53
- Tennessee Medicaid, 115, 307, 311–14
- terbutaline, 356, 357
- terfenadine, 7, 56, 94, 110
- terminal elimination phase, 48

- tetracycline, 198, 257, 273
thalidomide, 6, 143, 153, 175, 193, 231, 749, 751, 752, 753
theophylline, 56, 57
thiazide diuretics, 222, 256
thioridazine, 54
third-party administration (TPA) services, 296
Thomas v. Hoffman-LaRoche, Inc., 133
thrombocytopenic purpura/CCS, 223–4
thrombolytics, 114
ticrynafen, 6
timolol, 10
tililazad mesylate/aneurysmal subarachnoid hemorrhage, 595–7
tolcapone, 110
tolmetin, 37, 65, 66, 647
torsades de pointes, 291
tort theory, 126
toxic epidermal necrolysis/CCS, 224
toxic shock syndrome, 68, 69
Trade-Related Intellectual Property Rights (TRIPS), 80–3, 438
tramadol, 7, 242
Treatment IND (Investigational New Drug) Studies, 449
Treaty of Rome, 84
triazolam, 7
tricyclic antidepressants, 47
trimethoprim, 186
Triple Market Prenatal Screening Database (MSAFP), 268
troglitazone, 7, 110
trovafloxacin, 110
Tumour Registry, 268
type A reaction, 4, 68
type B reaction, 4, 68–9
type I error, 32
type II error, 32
- UN National Commission for the Protection of Human Subjects, 419
unanticipated beneficial effects, 554
unanticipated harmful effects, 553
uncertainty principle, 546
under the counter (UTC) market, 80
Unicef, 83
unit of analysis, 488
UnitedHealth Group, 115, 295–305, 407
 diagnosis and hospitalization data, 696
 feasibility studies
 future, 303–4
 overview, 295–6
 research databases, 296–301, 406
 validity of drug data in, 692
Universal Medical Device Nomenclature System, 739
University of Florida–Shands Teaching Hospital, 6
University of Washington, 248
university scientist, role in pharmacoepidemiology, 87
Upjohn Pharmaceuticals, 393
Uppsala Monitoring Centre in Sweden, 177, 180, 182, 189
Uruguay Round, 81
- Vaccine Adverse Event Report System (VAERS) (US), 157, 711, 713–15, 716, 719–20
Vaccine Associated Adverse Event (VAAE) (Canada), 711, 715
vaccine associated paralytic polio (VAPP), 708
Vaccine Identification Standards Initiative (VISI), 724
vaccine induced events, 715–16
Vaccine Injury Compensation Program, 723
Vaccine Injury Table, 709
vaccine potentiated events, 716
vaccine safety, 707–30
 analyses, confounding, and bias, 712
 assessment of causality, 711
 automated large-linked databases, 721–3
 clinical issues, 710
 clinical trials, 720–1
 exposure, 711
 future, 725–6
 logistical approaches, 712–23
 methodological approaches, 723–5
 outcome, 711–12
 policy issues, 708–10
 signal generation, 710
 spontaneous reporting system, 712–20
 assessment of causality, 716–18
 classifications, case definitions, and evaluative protocols, 715–16
 lessons learned to date, 719–20
 national experiences, 715
 signal detection, 718–19
 US experience, 713–14
Vaccine Safety Datalink (VSD), 254, 255, 279, 721–3
validity of drug and diagnosis data, 661–705
 computerized databases containing administrative data, 690–8
 clinical problems, 661–2
 etiologic studies, 668–9
 future, 698–700
 measurement error and point estimate of association, 667
 measures of association for measurement error, 667–8
 methodologic problems, 662–8
 methodologic studies, 670
 quantitative indices of measurement error, 664–7
 reliability, 666–7
 self-reported diagnosis/hospitalization data from *ad hoc* studies, 681–90, 682–7
 self-reported drug data from *ad hoc* questionnaire studies, 670–81
 accuracy, 67, 679–800
 medication use in pregnancy, 674–9
 oral contraceptive use, 670, 671
 postmenopausal hormone use, 673–4
 use of nonhormonal medications, 674, 675–7
validity of instrument, 605
valproate, 339, 751
VAMP Health, 375–6
vasectomy/prostatic cancer, 215, 222
VEN classification, 442
verapamil, 45, 56
 presystemic clearance, 55
Veterans' Administration Classification, 473
Veterans Administration Registry of Pacemaker Leads (US), 743
Viagra (sildenafil), 7, 47, 256
vigabatrin, 242–3
Virginia Mason Medical Center, 249
Vital Statistics Database, 269
vitamin B12, 86
vitamin K, 7, 56, 95
Viterbo v. The Dow Chemical Co., 133
- Walter Reed Army Medical Center HIV Staging System, 589
warfarin, 50, 57, 198
Washington Basic Health Plan, 249
Waxman-Hatch Act (1984) (US), 564
Welfare Reform Act (1996), 307, 308
Wells v. Ortho Pharmaceutical Corp., 145
Wilcoxon rank-sum test, 591
Wilcoxon signed rank test, 620

- willingness-to-pay analysis, 580
- Wisconsin Medicaid DUR program, 514
- withdrawals, **166**
- within-subject designs, 789–90
- Wooderson case, 145
- World Bank, 80, 83
- World Health Organization, 6, 83, 84, 165, 438
 - Action Program on Essential Drugs (ED), 80
 - Collaborating Center for International Drug Monitoring (Uppsala), 180, 715
 - Council for International Organizations of Medical Sciences (CIOMS), 450
 - Drug Utilization Research Group, 348, 468
 - Essential Drugs and Other Medicines (WHO/EDM), 434
 - International Drug Monitoring Program, 175–6, 473
 - Model Drugs Law, 441
- World Trade Organization, 80
 - Trade-Related Intellectual Property Rights (TRIPS) Agreement, 80–3, 438
- Xylocaine, 130
- “yellow card” system (UK), 238, 715
- zidovudine, 10, 64, 83
- zimeldine, 187–8
- zomepirac, 7, 66, 315
- Z-statistics, 64

Index compiled by Annette Musker

Part I

INTRODUCTION

1

What is Pharmacoepidemiology?

BRIAN L. STROM

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

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

In recent decades, modern medicine has been blessed with a pharmaceutical armamentarium that is much more powerful than what it had before. Although this has given us the ability to provide much better medical care for our 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 indicate that 100 000 Americans die each year from adverse drug reactions (ADRs), and 1.5 million US hospitalizations each year result from ADRs; yet, 20–70% of ADRs may be preventable.¹ The harm that drugs can cause has also led to the development of the field of pharmacoepidemiology, which is the focus of this book. More recently, the field has begun to expand its focus to include many issues other than adverse reactions, as well.

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

DEFINITION OF PHARMACOEPIDEMIOLOGY

Pharmacoepidemiology is the study of the use of and the effects of drugs in large numbers of people. The term pharmacoepidemiology obviously contains two components: “pharmaco” and “epidemiology.” In order to better appreciate and understand what is and what is not included in this new field, it is useful to compare its scope to

that of other related fields. The scope of pharmacoepidemiology will first be compared to that of clinical pharmacology, and then to that of epidemiology.

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 his infection, it is not sufficient to know that gentamicin has a small probability of causing renal disease. A good clinician should realize that a patient who has impaired liver function is at a greater risk of suffering from this adverse effect than one with normal liver function.² Pharmacoepidemiology can be useful in providing information about the beneficial and harmful effects of any drug, thus permitting a better assessment of the risk/benefit balance for the use of any particular drug in any particular patient.

Clinical pharmacology is traditionally divided into two basic areas: pharmacokinetics and pharmacodynamics. *Pharmacokinetics* is the study of the relationship between the dose administered of a drug and the serum or blood level achieved. It deals with drug absorption, distribution, metabolism, and excretion. *Pharmacodynamics* is the study of the relationship between drug level and drug effect. Together, these two fields allow one to

predict the effect one might observe in a patient from administering a certain drug regimen. Pharmacoepidemiology encompasses elements of both of these fields, exploring the effects achieved by administering a drug regimen. It does not normally involve or require the measurement of drug levels. However, pharmacoepidemiology can be used to shed light on the pharmacokinetics of a drug, such as exploring whether aminophylline is more likely to cause nausea when administered to a patient simultaneously taking cimetidine. However, 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*.³ Type A reactions tend to be common, dose-related, predictable, and less serious. They can usually be treated by simply reducing the dose of the drug. They tend to occur in individuals who have one of three characteristics. First, the individuals may have received more of a drug than is customarily required. Second, they may have received a conventional amount of the drug, but they may metabolize or excrete the drug unusually slowly, leading to drug levels that are too high. 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 most difficult to predict or even detect, and represent the major focus of pharmacoepidemiologic 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 10 and 11). However, determining causation in case reports of adverse reactions can be problematic (see Chapter 32), as can attempts to compare the effects of drugs in the same class. This has led academic investigators, industry, 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 judgement 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.

PHARMACOEPIDEMIOLOGY 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, that is 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,⁴ the major application of these principles is after drug marketing. This has primarily been in the context of postmarketing drug surveillance, although in recent years the interests of pharmacoepidemiologists have broadened considerably.

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.⁵ In response, the 1938 Food, Drug, and Cosmetic Act was passed. Preclinical toxicity testing was required for the first time. In addition, manufacturers were required to gather clinical data about drug safety and to submit these data to FDA before drug marketing. The FDA had 60 days to object to marketing or else it would proceed. No proof of efficacy was required.

Little attention was paid to adverse drug reactions until the early 1950s, when it was discovered that chloramphenicol could cause aplastic anemia.⁶ In 1952, the first textbook of adverse drug reactions was published.⁷ In the same

year, the AMA Council on Pharmacy and Chemistry established the first official registry of adverse drug effects, to collect cases of drug induced blood dyscrasias.⁸ In 1960, the FDA began to collect reports of adverse drug reactions and sponsored new hospital-based drug monitoring programs. The Johns Hopkins Hospital and the Boston Collaborative Drug Surveillance Program developed the use of in-hospital monitors to perform cohort studies to explore the short-term effects of drugs used in hospitals^{9,10} (see Chapter 12). This approach was later to be transported to the University of Florida–Shands Teaching Hospital, as well.¹¹

In the winter of 1961, the world experienced the infamous “thalidomide disaster.” Thalidomide was marketed as a mild hypnotic, and had no obvious advantage over other drugs in its class. Shortly after its marketing, a dramatic increase was seen in the frequency of a previously rare birth defect, phocomelia—the absence of limbs or parts of limbs, sometimes with the presence instead of flippers.¹² Epidemiologic studies established its cause to be *in utero* exposure to thalidomide. In the United Kingdom, this resulted in the establishment in 1968 of the Committee on Safety of Medicines. Later, the World Health Organization established a bureau to collect and collate information from this and other similar national drug monitoring organizations (see Chapter 11).

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 man. The data from these studies were required to be submitted to 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 DESI (Drug Efficacy Study Implementation) process, conducted by the National Academy of Sciences’ National Research Council with support from a contract from FDA, was not completed until 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.¹³ 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.^{14–18} These studies provided the first descriptive information on how physicians use drugs, and began a series of investigations of the frequency and determinants of poor prescribing (see also Chapters 29–31).

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, and especially the 1980s and 1990s 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 1970.¹⁹ 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.²⁰ The mid-1970s saw the discovery of the oculomucocutaneous syndrome caused by practolol, five years after drug marketing.²¹ In 1980, the drug ticrynafen was noted to cause

deaths from liver disease.²² In 1982, benoxaprofen was noted to do the same.²³ Subsequently the use of zomepirac, another nonsteroidal anti-inflammatory drug, was noted to be associated with an increased risk of anaphylactoid reactions.²⁴ Serious blood dyscrasias were linked to phenylbutazone.²⁵ Small intestinal perforations were noted to be caused by a particular slow release formulation of indomethacin.²⁶ 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 claim²⁷, (and see Chapters 9 and 42). Acute flank pain and reversible acute renal failure were noted to be caused by suprofen.²⁸ Isotretinoin was almost removed from the US market because of the birth defects it causes.^{29,30} The Eosinophilia–Myalgia syndrome was linked to a particular brand of L-tryptophan.³¹ Triazolam, thought by The Netherlands in 1979 to be subject to a disproportionate number of central nervous system side effects,³² was discovered by the rest of the world to be problematic in the early 1990s.^{33–35} Silicone breast implants, inserted by the millions in the US for cosmetic purposes, were accused of causing cancer, rheumatologic disease, and many other problems, and restricted from use except for breast reconstruction after mastectomy.³⁶ 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.^{37–41} 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.^{42,43} An epidemic of deaths from asthma in New Zealand was traced to fenoterol,^{44–46} and later data suggested that similar, although smaller, risks might be present with other β -agonist inhalers.⁴⁷ The possibility was raised of cancer from depot-medroxyprogesterone, resulting in initial refusal to allow its marketing for the purpose of contraception in the US,⁴⁸ multiple studies,^{49,50} and ultimate approval. Arrhythmias were linked to the use of the antihistamines terfenadine and astemizole.^{51,52} Hypertension, seizures, and strokes were noted from postpartum use

of bromocriptine.^{53,54} Multiple different adverse reactions were linked to temafloxacin.⁵⁵ 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.

Since the last edition of this book, drug crises have occurred due to allegations of liver toxicity from amoxicillin–clavulanic acid;⁵⁶ liver toxicity from bromfenac;^{57,58} cancer, myocardial infarction, and gastrointestinal bleeding from calcium channel blockers;^{59–67} arrhythmias with cisapride interactions;^{68,69} primary pulmonary hypertension and cardiac valvular disease from dexfenfluramine and fenfluramine;^{70,71} gastrointestinal bleeding, postoperative bleeding, deaths, and many other adverse reactions associated with ketorolac;^{72–75} multiple drug interactions with mibefradil;⁷⁶ thrombosis from newer oral contraceptives;^{77–80} myocardial infarction from sildenafil;⁸¹ seizures with tramadol;^{82,83} anaphylactic reactions from vitamin K;⁸⁴ liver toxicity from troglitazone;^{85–88} and intussusception from rotavirus vaccine.⁸⁹ Thirteen different prescription drug products have been removed from the US market, since 1980 alone (see Chapter 8).

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.

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 Unit, which extended the hospital based approach of the Boston Collaborative Drug Surveillance Program (Chapter 12) by collecting lifetime drug exposure histories from hospitalized patients and using these to perform hospital-based case–control studies⁹⁰ (see Chapter 13). 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.⁹¹ The Computerized Online Medicaid Analysis and Surveillance System was first developed in 1977, using Medicaid billing data to perform pharmacoepidemiologic studies⁹² (see Chapter 19). 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 Monitoring⁹³ (see Chapter 14). Each of these represented major contributions to the field of pharmacoepidemiology. A number of additional resources have been developed more recently, and are discussed in Part III of this book, along with the more established resources.

The 1980s and especially the 1990s 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 Part IV 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,⁹⁴ and the International Society for Pharmacoepidemiology issued, in 1996, guidelines for good epidemiology practices for drug, device, and vaccine research in the United States.⁹⁵ In the late 1990s, pharmacoepidemiology research has been increasingly hampered by concerns about patient confidentiality⁹⁶⁻¹⁰⁰ (see also Chapter 26).

Organizationally, in the US, the Prescription Drug User Fee Act of 1992 allowed the US FDA to charge manufacturers a fee for reviewing New Drug Applications. This provided additional resources to 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 FDA in the 1990s. Each drug costs \$300–500 million to develop; drug development will cost the pharmaceutical industry a total of \$24 billion in 1999. Yet, funds from the Prescription Drug User Fee Act of 1992 are prohibited from being used for drug safety regulation. In 1998, whereas 1400 FDA employees worked with the drug approval process, only 52 monitored safety; FDA spent only \$2.4 million in extramural safety research. This has coincided with the growing numbers of drug crises cited above. As another measure of drug safety problems, the FDA's new MedWatch program of collecting spontaneous reports of adverse reactions (see Chapter 10) 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.

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 FDA and other regulatory bodies are devoted to discovering rare unknown risks from new drugs. In response, there is growing concern, in Congress and the US public at least, that perhaps FDA is now approving drugs too *fast*.¹⁰¹ There are also calls for the development of an independent drug safety board, analogous to the National Transportation Safety Board,^{102,103} with a mission much wider than FDA's regulatory mission, to complement the latter. For example, such a board could investigate drug safety crises such as those cited above, looking for ways to prevent them, and could deal with issues such as improper physician use of drugs, the need for training, and the development of new approaches to the field of pharmacoepidemiology.

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 in humans 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 which 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. At the other extreme, a total of 500 patients would allow one to be 95% certain of detecting any adverse reactions which occur in six or more patients out of every 1000 exposed. Adverse reactions 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 PHARMACOEPIDEMOLOGY

The potential contributions of pharmacoepidemiology are only beginning to be realized, as the field is 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.¹⁰⁴ In this section of this chapter, we will first review the potential for pharmacoepidemiologic studies to supplement the information available prior to marketing, and then review the new types of information obtainable from postmarketing pharmacoepidemiologic studies but not obtainable prior to drug marketing. Finally, we will review the general, and probably most important, potential contributions such studies can make. In each case, the relevant information available from premarketing studies will be briefly examined first, to clarify how postmarketing studies can supplement this information.

SUPPLEMENTARY INFORMATION

Premarketing studies of drug effects are necessarily limited in size. After marketing, nonexperimental

Table 1.1. Potential contributions of pharmacoepidemiology

-
- | | |
|-----|---|
| (A) | Information which supplements that 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, in 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 |
-

epidemiologic studies can be performed, evaluating the effects of drugs administered as part of ongoing medical care. These allow the cost-effective accumulation of much larger numbers of patients than those studied prior to marketing, resulting in a more precise measurement of the incidence of adverse and beneficial drug effects (see Chapter 3). For example, at the time of drug marketing, prazosin was known to cause a dose dependent first dose syncope,^{105,106} but the FDA requested the manufacturer to conduct a post-marketing surveillance study in the US to quantify its incidence more precisely.⁹¹ 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.^{107,108}

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.¹⁰⁹

Additionally, for reasons of statistical efficiency, premarketing clinical trials generally seek subjects who are as homogeneous as possible, in order to reduce unexplained variability in the outcome variables measured and increase the probability of detecting a difference between the study groups, if one truly exists. For these reasons, certain patients are often excluded, including those with other illnesses or those who are receiving other drugs. Postmarketing studies can explore how factors such as other illnesses and other drugs might modify the effects of the drugs, as well as looking at the effects of differences in drug regimen, compliance, etc.¹¹⁰ 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.¹¹¹

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.¹¹²

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.²⁰

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 Chapters 29 and 30).

In most cases, premarketing studies are performed using selected patients who are closely observed. Rarely are there any significant overdoses in this population. Thus, the study of the effects of a drug when ingested in extremely high doses is rarely possible before drug marketing. Again, this must await postmarketing pharmacoepidemiologic studies.¹¹³

Finally, it is only recently that our society has become more sensitive to the costs of medical care, and only recently have the techniques of health economics been applied to evaluate the cost implications of drug use.¹¹⁴ It is clear that the exploration of the costs of drug use requires consideration of much more than just the costs of the drugs themselves. The costs of a drug's adverse effects may be much more than the costs of the drug, if these adverse effects result in additional medical care and possibly even hospitalizations.¹¹⁵ 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 35).

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.

REFERENCES

1. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *J Am Med Assoc* 1998; **279**: 1200–5.
2. Moore RD, Smith CR, Lietman PS. Increased risk of renal dysfunction due to interaction of liver disease and aminoglycosides. *Am J Med* 1986; **80**: 1093–7.
3. Rawlins MD, Thompson JW. Pathogenesis of adverse drug reactions. In: Davies DM, ed., *Textbook of Adverse Drug Reactions*. Oxford: Oxford University Press, 1977; 44.
4. Strom BL. Integrating pharmacoepidemiology into the design and conduct of clinical trials. *J Clin Res Drug Dev* 1988; **2**: 161–8.
5. Geiling EMK, Cannon PR. Pathogenic effects of elixir of sulfanilimide (diethylene glycol) poisoning. *J Am Med Assoc* 1938; **111**: 919–26.
6. Wallerstein RO, Condit PK, Kasper CK, Brown JW, Morrison FR. Statewide study of chloramphenicol therapy and fatal aplastic anemia. *J Am Med Assoc* 1969; **208**: 2045–50.
7. Meyler L. *Side Effects of Drugs*. Amsterdam: Elsevier, 1952.
8. Erslev AJ, Wintrobe MM. Detection and prevention of drug induced blood dyscrasias. *J Am Med Assoc* 1962; **181**: 114–9.
9. Cluff LE, Thornton GF, Seidl LG. Studies on the epidemiology of adverse drug reactions. I. Methods of surveillance. *J Am Med Assoc* 1964; **188**: 976–83.
10. Miller RR, Greenblatt DJ. *Drug Effects in Hospitalized Patients*. New York: Wiley, 1976.
11. Caranasos GJ, Stewart RB, Cluff LE. Drug-induced illness leading to hospitalization. *J Am Med Assoc* 1974; **228**: 713–7.
12. Lenz W. Malformations caused by drugs in pregnancy. *Am J Dis Child* 1966; **112**: 99–106.

13. Wardell WM. Therapeutic implications of the drug lag. *Clin Pharmacol Ther* 1974; **15**: 73–96.
14. Lee JAH, Draper PA, Weatherall M. Prescribing in three English towns. *Milbank Mem Fund Q* 1965; **43**: 285–90.
15. Muller C. Medical review of prescribing. *J Chron Dis* 1965; **18**: 689–96.
16. Meade TW. Prescribing of chloramphenicol in general practice. *Br Med J* 1967; **1**: 671–4.
17. Joyce CRB, Last JM, Weatherall M. Personal factors as a cause of differences in prescribing by general practitioners. *Br J Prev Soc Med* 1968; **22**: 170–7.
18. Muller C. Outpatient drug prescribing related to clinic utilization in four New York City hospitals. *Health Serv Res* 1968; **3**: 142–54.
19. Kono R. Trends and lessons of SMON research. In: Soda T (ed). *Drug-Induced Sufferings*. Princeton, NJ: Excerpta Medica, 1980; 11.
20. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *New Engl J Med* 1971; **284**: 878–81.
21. Wright P. Untoward effects associated with pralolol administration. Oculomucocutaneous syndrome. *Br Med J* 1975; **1**: 595–8.
22. Anonymous. Ticrynafen recalled. *FDA Drug Bull* 1980; **10**: 3–4.
23. Suspension of benoxaprofen (Opren). *Br Med J* 1982; **285**: 519–20.
24. 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.
25. Inman WHW. Study of fatal bone marrow depression with special reference to phenylbutazone and oxyphenbutazone. *Br Med J* 1977; **1**: 1500–5.
26. Committee on Safety of Medicines. *Osmosin (Controlled Release Indomethacin)*. (Current Problems No. 11). London: Committee on Safety of Medicines, 1983.
27. Barash CI, Lasagna L. The Bendectin saga: “voluntary” discontinuation. *J Clin Res Drug Dev* 1987; **1**: 277–92.
28. 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.
29. Marwick C. FDA ponders approaches to curbing adverse effects of drug used against cystic acne. *J Am Med Assoc* 1988; **259**: 3225–30.
30. Stern RS. When a uniquely effective drug is teratogenic: the case of isotretinoin. *New Engl J Med* 1989; **320**: 1007–9.
31. Hertzman PA, Falk H, Kilbourne EM, Page S, Shulman LE. The Eosinophilia–Myalgia Syndrome: The Los Alamos Conference. *J Rheumatol* 1991; **18**: 867–73.
32. Meyboom RHB. The triazolam experience in 1979 in The Netherlands, a problem of signal generation and verification. In: Strom BL, Velo GP, eds, *Drug Epidemiology and Postmarketing Drug Surveillance*. New York: Plenum, 1992; 159–67.
33. Triazolam’s status in European Community. *Lancet* 1991; **338**: 1586–7.
34. Brahams D. Triazolam suspended. *Lancet* 1991; **338**: 938.
35. Triazolam hearing. *Lancet* 1992; **339**: 1291.
36. Randall T. Penile, testicular, other silicone implants soon will undergo FDA review [news]. *J Am Med Assoc* 1992; **267**: 2578–9.
37. Egger M, Smith GD, Imhoof H, Teuscher A. Risk of severe hypoglycaemia in insulin treated diabetic patients transferred to human insulin: a case control study. *Br Med J* 1991; **303**: 617–21.
38. Egger M, Smith GD, Teuscher AU, Teuscher A. Influence of human insulin on symptoms and awareness of hypoglycaemia: a randomised double blind crossover trial. *Br Med J* 1991; **303**: 622–6.
39. Colagiuri S, Miller JJ, Petocz P. Double-blind crossover comparison of human and porcine insulins in patients reporting lack of hypoglycaemia awareness. *Lancet* 1992; **339**: 1432–5.
40. Jick H, Hall GC, Dean AD, Jick SS, Derby LE. A comparison of the risk of hypoglycemia between users of human and animal insulins. 1. Experience in the United Kingdom. *Pharmacotherapy* 1990; **10**: 395–7.
41. Jick SS, Derby LE, Gross KM, Jick H. Hospitalizations because of hypoglycemia in users of animal and human insulins. 2. Experience in the United States. *Pharmacotherapy* 1990; **10**: 398–9.
42. Teicher MH, Glod C, Cole JO. Emergence of intense suicidal preoccupation during fluoxetine treatment. *Am J Psychiatry* 1990; **147**: 207–10.
43. Beasley CM, Dornseif BE, Bosomworth JC, Sayler ME, Rampey AH, Heiligenstein JH, Thompson VL, Murphy DJ, Masica DN. Fluoxetine and suicide: a meta-analysis of controlled trials of treatment for depression. *Br Med J* 1991; **303**: 685–92.
44. Crane J, Pearce N, Flatt A, Burgess C, Jackson R, Kwong T, Ball M, Beasley R. Prescribed fenoterol and death from asthma in New Zealand, 1981–83: case–control study. *Lancet* 1989; **1**: 917–22.
45. Pearce N, Grainger J, Atkinson M, Crane J, Burgess C, Culling C, Windom H, Beasley R. Case–control study of prescribed fenoterol and death from asthma in New Zealand, 1977–81. *Thorax* 1990; **45**: 170–5.

46. Grainger J, Woodman K, Pearce N, Crane J, Burgess C, Keane A, Beasley R. Prescribed fenoterol and death from asthma in New Zealand, 1981-7: a further case-control study. *Thorax* 1991; **46**:105-11.
47. Spitzer WO, Suissa S, Ernst P, Horwitz RI, Habbick B, Cockcroft D, Boivin JF, McNutt M, Buist AS, Rebeck AS. The use of beta-agonists and the risk of death and near death from asthma. *New Engl J Med* 1992; **326**: 501-6.
48. Rosenfield A, Maine D, Rochat R, Shelton J, Hatcher RA. The Food and Drug Administration and medroxyprogesterone acetate. What are the issues? *J Am Med Assoc* 1983; **249**: 2922-8.
49. WHO Collaborative Study of Neoplasia and Steroid Contraceptives. Breast cancer and depot-medroxyprogesterone acetate: a multinational study. *Lancet* 1991; **338**: 833-8.
50. WHO Collaborative Study of Neoplasia and Steroid Contraceptives. Depot-medroxyprogesterone acetate (DMPA) and risk of invasive squamous cell cervical cancer. *Contraception* 1992; **45**: 299-312.
51. Nightingale SL. From the Food and Drug Administration: Warnings issued on nonsedating antihistamines terfenadine and astemizole. *J Am Med Assoc* 1992; **268**: 705.
52. Ahmad SR. Antihistamines alert. *Lancet* 1992; **340**: 542.
53. Rothman KJ, Funch DP, Dreyer NA. Bromocriptine and puerperal seizures. *Epidemiology* 1990; **1**: 232-8.
54. Gross TP. Bromocriptine and puerperal seizures. *Epidemiology* 1991; **2**: 234-5.
55. Finch RG. The withdrawal of temafloxacin: Are there implications for other quinolones? *Drug Safety* 1993; **8**: 9-11.
56. Nathani MG, Mutchnick MG, Tynes DJ, Ehrnpreis MN. An unusual case of amoxicillin/clavulanic acid-related hepatotoxicity. *Am J Gastroenterol* 1998; **93**: 1363-5.
57. Hunter EB, Johnston PE, Tanner G, Pinson CW, Awad JA. Bromfenac (Duract)-associated hepatic failure requiring liver transplantation. *Am J Gastroenterol* 1999; **94**: 2299-301.
58. Moses PL, Schroeder B, Alkhatib O, Ferrentino N, Suppan T, Lidofsky SD. Severe hepatotoxicity associated with bromfenac sodium. *Am J Gastroenterol* 1999; **94**: 1393-6.
59. Furberg CD, Psaty BM, Meyer JV. Nifedipine. Dose-related increase in mortality in patients with coronary heart disease. *Circulation* 1995; **92**: 1326-31.
60. Pahor M, Guralnik JM, Furberg CD, Carbonin P, Havlik R. Risk of gastrointestinal haemorrhage with calcium antagonists in hypertensive persons over 67 years old. *Lancet* 1996; **347**: 1061-5.
61. Pahor M, Guralnik JM, Ferrucci L, Corti MC, Salive ME, Cerhan JR, Wallace RB, Havlik RJ. Calcium-channel blockade and incidence of cancer in aged populations. *Lancet* 1996; **348**: 493-7.
62. Braun S, Boyko V, Behar S, Reicher-Reiss, H, Shotan A, Schlesinger Z, Rosenfeld T, Palant A, Friedensohn A, Laniado S, Goldbourt U. Calcium antagonists and mortality in patients with coronary artery disease: a cohort study of 11, 575 patients. *J Am College Cardiol* 1996; **28**: 7-11.
63. Kostis JB, Lacy CR, Cosgrove NM, Wilson AC. Association of calcium channel blocker use with increased rate of acute myocardial infarction in patients with left ventricular dysfunction. *Am Heart J* 1997; **133**: 550-7.
64. Jick H, Jick S, Derby LE, Vasilakis C, Myers MW, Meier CR. Calcium-channel blockers and risk of cancer. *Lancet* 1997; **349**: 525-8.
65. Fitzpatrick AL, Daling JR, Furberg CD, Kronmal RA, Weissfeld JL. Use of calcium channel blockers and breast carcinoma risk in postmenopausal women. *Cancer* 1997; **80**: 1438-47.
66. Rosenberg L, Rao RS, Palmer JR, Strom BL, Stolley PD, Zauber AG, Warshauer ME, Shapiro S. Calcium channel blockers and the risk of cancer. *J Am Med Assoc* 1998; **279**: 1000-4.
67. Michels KB, Rosner BA, Walker AM, Stampfer MJ, Manson JE, Colditz GA, Hennekens CH, Willett WC. Calcium channel blockers, cancer incidence, and cancer mortality in a cohort of U.S. women: the Nurses' Health Study. *Cancer* 1998; **83**: 2003-7.
68. Trinkle R. Comment: syncopal episodes associated with cisapride. *Ann Pharmacother* 1999; **33**: 251.
69. Piquette RK. Torsade de pointes induced by cisapride/clarithromycin interaction. *Ann Pharmacother* 1999; **33**: 22-6.
70. Kramer MS, Lane DA. Aminorex, dexfenfluramine, and primary pulmonary hypertension. *J Clin Epidemiol* 1998; **51**: 361-4.
71. Connolly HM, Crary JL, McGoon MD, Hensrud DD, Edwards BS, Edwards WD, Schaff HV. Valvular heart disease associated with fenfluramine-phentermine. *New Engl J Med* 1997; **337**: 581-8.
72. Strom BL, Berlin JA, Kinman JL, Spitz PW, Hennessy S, Feldman H, Kimmel S, Carson JL. Parenteral ketorolac and risk of gastrointestinal and operative site bleeding: A postmarketing surveillance study. *J Am Med Assoc* 1996; **275**: 376-82.
73. Feldman HI, Kinman JL, Berlin JA, Hennessy S, Kimmel SE, Farrar J, Carson JL, Strom BL. Parenteral ketorolac: the risk for acute renal failure. *Ann Intern Med* 1997; **126**: 193-9.
74. Hennessy S, Kinman JL, Berlin JA, Feldman HI, Carson JL, Kimmel SE, Farrar J, Harb G, Strom BL. Lack of hepatotoxic effects of parenteral

- ketorolac in the hospital setting. *Arch Intern Med* 1997; **157**: 2510–4.
75. Garcia Rodriguez LA, Cattaruzzi C, Troncon MG, Agostinis L. Risk of hospitalization for upper gastrointestinal tract bleeding associated with ketorolac, other nonsteroidal anti-inflammatory drugs, calcium antagonists, and other antihypertensive drugs. *Arch Intern Med* 1998; **158**: 33–9.
 76. Mullins ME, Horowitz BZ, Linden DH, Smith GW, Norton RL, Stump J. Life-threatening interaction of mibefradil and beta-blockers with dihydropyridine calcium channel blockers. *J Am Med Assoc* 1998; **280**: 157–8.
 77. Spitzer WO, Lewis MA, Heinemann LA, Thorogood M, MacRae KD. Third generation oral contraceptives and risk of venous thromboembolic disorders: an international case-control study. *Br Med J* 1996; **312**: 83–8.
 78. Jick H, Jick SS, Gurewich V, Myers MW, Vasilakis C. Risk of idiopathic cardiovascular death and nonfatal venous thromboembolism in women using oral contraceptives with differing progestogen components. *Lancet* 1995; **346**: 1589–93.
 79. WHO Technical Report Series. *Cardiovascular Disease and Steroid Hormone Contraception*. Geneva: World Health Organization, 1998.
 80. de Bruijn SF, Stam J, Vandenbroucke JP. Increased risk of cerebral venous sinus thrombosis with third-generation oral contraceptives. *Lancet* 1998; **351**: 1404.
 81. Feenstra J, van Drie-Pierik RJ, Lacle CF, Stricker BH. Acute myocardial infarction associated with sildenafil. *Lancet* 1998; **352**: 957–8.
 82. Kahn LH, Alderfer RJ, Graham DJ. Seizures reported with tramadol. *J Am Med Assoc* 1997; **278**: 1661.
 83. Jick H, Derby LE, Vasilakis C, Fife D. The risk of seizures associated with tramadol. *Pharmacotherapy* 1998; **18**: 607–11.
 84. Pereira SP, Williams R. Adverse events associated with Vitamin K₁: results of a worldwide postmarketing surveillance programme. *Pharmacoepidemiol Drug Safety* 1998; **7**: 173–82.
 85. Misbin RI. Troglitazone-associated hepatic failure. *Ann Intern Med* 1999; **130**: 330.
 86. Vella A, de Groen PC, Dinneen SF. Fatal hepatotoxicity associated with troglitazone. *Ann Intern Med* 1998; **129**: 1080.
 87. Neuschwander-Tetri BA, Isley WL, Oki JC, Ramrakhiani S, Quason SG, Phillips NJ, Brunt EM. Troglitazone-induced hepatic failure leading to liver transplantation. A case report. *Ann Intern Med* 1998; **129**: 38–41.
 88. Gitlin N, Julie NL, Spurr CL, Lim KN, Juarbe HM. Two cases of severe clinical and histologic hepatotoxicity associated with troglitazone. *Ann Intern Med* 1998; **129**: 36–8.
 89. Anonymous. Intussusception among recipients of rotavirus vaccine—United States, 1998–1999. *Morbidity Mortality Weekly Rep* 1999; **48**: 577–81.
 90. Slone D, Shapiro S, Miettinen OS. Case-control surveillance of serious illnesses attributable to ambulatory drug use. In: Colombo F, Shapiro S, Slone D, Tognoni G, eds, *Epidemiological Evaluation of Drugs*. MA: PSG, 1977; 59–82.
 91. Joint Commission on Prescription Drug Use. *Final Report*. Washington: Joint Commission on Prescription Drug Use, 1980.
 92. Strom BL, Carson JL, Morse ML, Leroy AA. The Computerized Online Medicaid Analysis and Surveillance System: a new resource for postmarketing drug surveillance. *Clin Pharmacol Ther* 1985; **38**: 359–64.
 93. Inman WHW. Prescription Event Monitoring. *Acta Med Scand Suppl* 1984; **683**: 119–26.
 94. 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.
 95. Andrews EB, Avorn J, Bortnichak EA, Chen R, Dai WS, Dieck GS, Edlavitch S, Freiman J, Mitchell AA, Nelson RC, Neutel CI, Stergachis A, Strom BL, Walker AM. Guidelines for good epidemiology practices for drug, device, and vaccine research in the United States. *Pharmacoepidemiol Drug Safety* 1996; **5**: 333–8.
 96. Neutel CI. Privacy issues in research using record linkage. *Pharmacoepidemiol Drug Safety* 1997; **6**: 367–9.
 97. Melton LJ III. The threat to medical-records research. *New Engl J Med* 1997; **337**: 1466–70.
 98. Mann RD. Data privacy and confidentiality. *Pharmacoepidemiol Drug Safety* 1999; **8**: 245.
 99. Andrews EB. Data privacy, medical record confidentiality, and research in the interest of public health. *Pharmacoepidemiol Drug Safety* 1999; **8**: 247–60.
 100. Mann RD. The issue of data privacy and confidentiality in Europe—1998. *Pharmacoepidemiol Drug Safety* 1999; **8**: 261–4.
 101. Kleinke JD, Gottlieb S. Is the FDA approving drugs too fast? Probably not—but drug recalls have sparked debate. *Br Med J* 1998; **317**: 899.
 102. Wood AJ, Stein CM, Woosley R. Making medicines safer—the need for an independent drug safety board. *New Engl J Med* 1998; **339**: 1851–4.
 103. Moore TJ, Psaty BM, Furberg CD. Time to act on drug safety. *J Am Med Assoc* 1998; **279**: 1571–3.
 104. 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.
 105. Rosendorff C. Prazosin: severe side effects are dose-dependent. *Br Med J* 1976; **2**: 508.

106. Graham RM, Thornell IR, Gain JM, Bagnoli C, Oates HF, Stokes GS. Prazosin: the first-dose phenomenon. *Br Med J* 1976; **2**: 1293–4.
107. Kaitin KI. Case studies of expedited review: AZT and L-Dopa. *Law Med Health Care* 1991; **19**: 242–6.
108. Wastila LJ, Lasagna L. The history of zidovudine (AZT). *J Clin Res Pharmacoevidemiol* 1990; **4**: 25–37.
109. McKenzie MW, Marchall GL, Netzloff ML, Cluff LE. Adverse drug reactions leading to hospitalization in children. *JPediatr* 1976; **89**: 487–90.
110. May FE, Stewart RB, Cluff LE. Drug interactions and multiple drug administration. *Clin Pharmacol Ther* 1977; **22**: 322–8.
111. Nelson WL, Fraunfelder FT, Sills JM, Arrowsmith JB, Kuritsky JN. Adverse respiratory and cardiovascular events attributed to timolol ophthalmic solution, 1978–1985. *Am J Ophthalmol* 1986; **102**: 606–11.
112. The International Agranulocytosis and Aplastic Anemia Study. Risks of agranulocytosis and aplastic anemia: a first report of their relation to drug use with special reference to analgesics. *J Am Med Assoc* 1986; **256**: 1749–57.
113. Stewart RB, Forgnone M, May FE, Forbes J, Cluff LE. Epidemiology of acute drug intoxications: patient characteristics, drugs, and medical complications. *Clin Toxicol* 1974; **7**: 513–30.
114. Eisenberg JM. New drugs and clinical economics: analysis of cost-effectiveness analysis in the assessment of pharmaceutical innovations. *Rev Infect Dis* 1984; **6** (suppl. 4): S905–8.
115. Morse ML, Leroy AA, Gaylord TA, Kellenberger T. Reducing drug therapy-induced hospitalization: Impact of drug utilization review. *Drug Inf J* 1982; **16**: 199–202.

2

Study Designs Available for Pharmacoepidemiology Studies

BRIAN L. STROM

Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine,
Philadelphia, PA, 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 error one can make in designing a study. Next the chapter will review the “Criteria for the causal nature of an association,” which is how one can decide whether an association demonstrated in a particular study is, in fact, a causal association. Finally, the specific study designs available for epidemiologic studies, or in fact for any clinical studies, will be reviewed. The next chapter discusses a specific methodologic issue which needs to be addressed in any study, but which is of particular importance for 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.^{1–13} 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 studies a sample of study subjects. Second, one generalizes the information obtained in this sample of study subjects, drawing 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

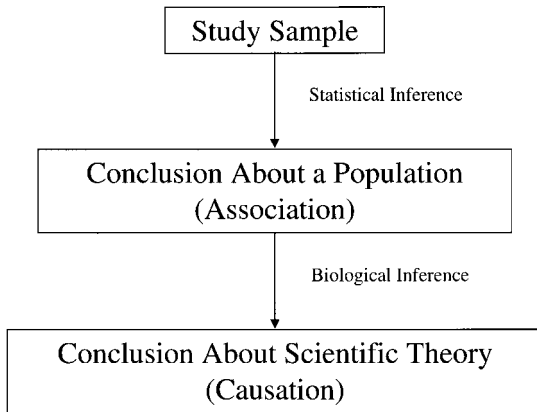


Figure 2.1. Overview of the scientific method.

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

At this point, one would be tempted to make a generalization that enalapril lowers blood pressure in middle aged hypertensive men. However, one must explore whether this observation could have occurred simply by chance, that is due to random variation. If this were true, then the observation might not have been made if one had chosen a sample of 40 different 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 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 judgement. To assist in making that judgement, 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 error that can be made in performing a study and the different types of association that each results in.

TYPES OF ERROR ONE CAN MAKE IN PERFORMING A STUDY

There are four basic types of association that can be observed in a study (Table 2.1). The basic purpose of research is to differentiate among them.

Table 2.1. Types of association between factors under study

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

First, of course, one could have no association.

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

The other possible mechanism for creating an artificial 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 bias.¹⁴ 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 an interviewer bias. Another example would be a study of drug-induced birth defects which compares children with birth defects to children without birth defects. A mother of an abnormal child 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 normal child, because of the unfortunate experience she has undergone. The improved recall in the mothers of the abnormal children may result in false apparent

associations between drug exposure and birth defects. This systematic difference in recall is referred to as a recall bias.¹⁵

Note that biases, once present, cannot be corrected. They represent errors in the study design which 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.

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, most examples of confounding are not. In designing a study, one must consider every variable which can be associated with the risk factor under study and 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.

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

Thus, there are three possible types of error that can be produced in a study: random error, bias,

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

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 in either the design of the study or in its analysis. If all three types of error 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,¹⁶ but have been described in various forms since, each with some modification. Probably the best known description of them was in the first Surgeon General’s Report on Smoking and Health,¹⁷ 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 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.¹⁸ 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.¹⁹ 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

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
-

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 gets infected with the measles virus develops clinical measles. Certainly, not everyone who smokes develops lung cancer, and not everyone who develops lung cancer was a smoker. This is one of the major points the tobacco industry stresses when it attempts to make the claim that cigarette smoking has not been proven to cause lung cancer. Some authors even omit this as a criterion, as it is so rarely met. When it is met, however, it provides extremely strong support for a conclusion that an association is causal.

The fifth criterion listed in Table 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 its size. To evaluate this, one asks whether the magnitude of the 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 30.0.¹⁹

A dose–response relationship is an extremely important and commonly used concept in clinical

pharmacology and is used similarly in epidemiology. A *dose–response relationship* exists when an increase in the intensity of an exposure results in an increased risk of the disease under study. Equivalent to this is a *duration–response relationship*, which exists when a longer exposure causes an increased risk of the disease. The presence of either a dose–response relationship or a duration–response relationship strongly implies that an association is, in fact, a causal association. Certainly in the example of cigarette smoking and lung cancer, it has been shown repeatedly that an increase in either the number of cigarettes smoked each day or in the number of years of smoking increases the risk of developing lung cancer.¹⁹

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, or 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 goes from the designs at the bottom of the table to those at the top of the table, studies get progressively harder to perform, but are progressively more convincing. In other words, associations shown by studies using designs at the top of the list are more likely to be causal associations than associations shown by studies using designs at the bottom of the list. The association between cigarette smoking and lung cancer has been reproduced in multiple well designed studies, using analyses of secular trends, case–control studies, and cohort studies. However, it has not been shown using a randomized clinical trial, which is the “Cadillac” of study designs, as will be discussed below. This is the other major defense used by the tobacco industry. Of course, it would not be ethical nor logistically feasible to randomly allocate individuals to smoke or not to smoke and expect them to follow that for 20 years.

The issue of causation is discussed more in Chapter 9 as it relates to litigation, in Chapters 10 and 11 as it relates to the process of spontaneous

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 testing

reporting of adverse drug reactions, and in Chapter 32 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 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 which has become apparent in the relatively recent past is clear cell vaginal adenocarcinoma occurring in young women exposed *in utero* to diethylstilbestrol.²⁰ 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 which 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 will become comatose again, and then if he or she is given another dose of naloxone the patient will awaken again. This, especially if repeated a few times, would represent strong evidence that the drug is indeed effective as a narcotic antagonist. This type of challenge–rechallenge situation is relatively uncommon, however, as physicians generally will avoid exposing a patient to a drug if the patient experienced an adverse reaction to it in the past. This issue is discussed in more detail in Chapters 10, 11, and 32.

CASE SERIES

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

After drug marketing, case series are most useful for two related purposes. First, they can be useful for quantifying the incidence of an adverse reaction. Second, they can be useful for being certain that any particular adverse effect of concern does not occur in a population which is larger than that studied prior to drug marketing. The so-called “Phase IV” postmarketing surveillance study of prazosin was conducted for the former reason, to quantitate the incidence of first dose syncope from prazosin.²¹ The “Phase IV” postmarketing surveillance study of cimetidine²² 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.²¹ In both examples, the manufacturer asked its sales representatives to recruit physicians to participate in the study. Each participating physician then enrolled the next series of patients for whom the drug was prescribed.

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

ANALYSES OF SECULAR TRENDS

Analyses of secular trends, also called “ecological studies,” examine trends in an exposure that is a presumed cause and trends in a disease that is a presumed effect and test whether the trends coincide. These trends can be examined over time or across geographic boundaries. In other words, one could analyze data from a single region and examine how it 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.²³

Analyses of secular trends are useful for rapidly providing evidence for or against an hypothesis. However, these studies lack data on individuals; they only study groups. 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.²⁴ This is certainly consistent with the increasing rates of cigarette smoking observed in women until the mid-1960s,²⁵ 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 which 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. A number of such studies have been performed, generally demonstrating a strong association between the use of oral contraceptives and venous thromboembolism.²⁶

Case-control studies can be particularly useful when one wants to study multiple possible causes of a single disease, as one can study any number of

exposures as potential risk factors using the same cases and controls. 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 eight cases and 40 controls,²⁰ rather than the many thousands of exposed subjects that would have been required for a cohort study of this question.

Case-control studies generally obtain their information on exposures retrospectively, that is recreating events that happened in the past. Information is generally obtained by abstracting medical records or by administering questionnaires or interviews. As such, they are subject to limitations in the validity of retrospectively collected exposure information. Also, the proper selection of controls can be a challenging task, and appropriate control selection can lead to a selection bias, which may lead to incorrect conclusions. Nevertheless, when case-control studies are well done, 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 one for pharmacoepidemiology studies.

COHORT STUDIES

Cohort studies are studies which 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 was noted using analyses of secular trends and case-control studies.^{27,28} Cohort studies can be

performed either prospectively, that is simultaneous with the events under study, or retrospectively, that is after the events under study, 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 on the basis of the presence or absence of a disease, and their antecedent exposures are then studied. Patients are recruited into cohort studies on the basis of the presence or absence of an exposure, and their subsequent disease course is then studied.

Cohort studies have the major advantage of being free of the major problem that plagues case-control studies: the difficult process of selecting an undiseased control group. In addition, prospective cohort studies are free of the problem of the questionable validity of retrospectively collected data. For these reasons, an association demonstrated by a cohort study is more likely to be a causal association than one demonstrated by a case-control study. Furthermore, cohort studies are particularly useful when one is studying multiple possible outcomes from a single exposure, especially a relatively uncommon exposure. Thus, they are 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 statistic reported from these studies is the relative risk. The *relative risk* is the ratio of the incidence rate of an outcome in the exposed group to the incidence rate of the outcome in the unexposed group. A relative risk of greater than 1.0 means that exposed subjects have a *greater* risk of the disease under study than unexposed subjects, or that the exposure appears to cause the disease. A relative risk less than 1.0 means that exposed subjects have a *lower* risk of the disease than unexposed subjects, or that the exposure seems to protect against the disease. A relative risk of 1.0 means that exposed subjects and unexposed subjects have the same risk of developing the disease, or that the exposure and the disease appear unrelated.

One can calculate a relative risk directly from the results of a cohort study. However, in a case-control study one cannot determine the size of either the exposed population or the unexposed population that the diseased subjects and undiseased subjects 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 an estimate of the relative risk when the disease under study is relatively rare. Since case-control studies are generally used to study rare diseases, there generally is very close agreement between the odds ratio and the relative risk, and the results of case-control studies are often loosely referred to as relative risks, although they are in fact odds ratios.

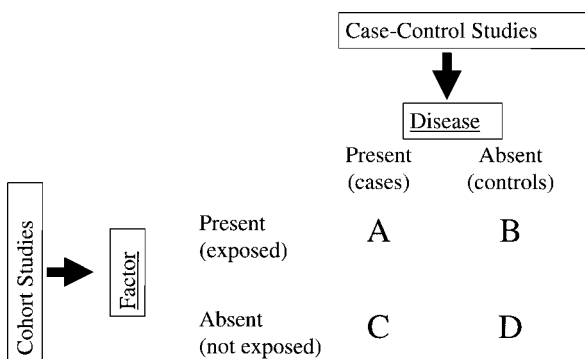


Figure 2.2. Cohort and case-control studies provide similar information, but approach data collection from opposite directions. (Reprinted with permission from Strom BL. Medical databases in post-marketing drug surveillance. *Trends in Pharmaceutical Sciences* 1986; 7: 377.)

Both relative risks and odds ratios can be reported with *p-values*. These *p-values* allow one to determine whether 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.

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. A 95% confidence interval around a relative risk means that we can be 95% confident that the true relative risk lies in this range. If a 95% confidence interval around a relative risk excludes 1.0, then the finding is statistically significant with a *p-value* of less than 0.05. A confidence interval provides much more information than a *p-value* however. As an example, a study which 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 which 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 which yields a relative risk (95% confidence interval) of 10.0 (9.8–10.2) precisely quantifies a tenfold increase in risk. A study which 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 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.²⁶ However, the risk of myocardial infarction in nonsmoking women in their 20s is so low that even a fivefold increase in that risk would still not be of public health importance. In contrast, women in their 40s are at higher risk, especially if they are cigarette smokers as well. Thus, oral contraceptives should not be used in these women.²⁶

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

RANDOMIZED CLINICAL TRIALS

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

However, even randomized clinical trials are not without their problems. The randomized clinical trial described 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.²⁹ However, they remain the "gold standard" by which the other designs must be judged.

DISCUSSION

Thus, a series of different study designs are available (Table 2.4), each with its own 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. These are all studies in which the investigator does not control the therapy, but simply observes and evaluates the results of ongoing medical care. Case reports, case series, and analyses of secular trends have also been referred to as *descriptive studies*. Case-control studies, cohort studies, and randomized clinical trials all have control groups, and have been referred to as *analytic studies*. The analytic study designs can be classified in two major ways, by how subjects are selected into the study and by how data are collected for the study (see Table 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 on the basis of the presence or absence of a disease, while cohort studies select subjects into the study on the basis of 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

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

collected *prospectively*, that is simultaneously with the events under study, or *retrospectively*, that is after the events under study. 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. These study the situation at one point in time only. 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 could 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 a number 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 used the term "prospective study" to refer to a cohort study, confusing the two concepts inherent in the classification schemes presented in the table. Other authors use the term "retrospective study" to refer to any 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.

CONCLUSIONS

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 upwards, from case reports and case series suggesting an association, to analyses of trends and case-control studies exploring them. Finally, if a study question warrants the delay and investment, cohort studies and randomized clinical trials can be performed.

For example, regarding the question of whether oral contraceptives cause venous thromboembolism, an association was first suggested by case reports and case series, then explored in more detail by analyses of trends and a series of case-control studies.²⁶ 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.^{27,28} This question might even be worth the investment of a randomized clinical trial, except that it would not be feasible nor 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 studies³⁰ and analyses of secular trends.³¹ 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.³²

In general, however, clinical, regulatory, commercial, and legal decisions need to be made on the basis of the best evidence available at the time of the decision. To quote Sir Austin Bradford Hill,¹⁶

All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.

Who knows, asked Robert Browning, but the world may end tonight? True, but on available evidence most of us make ready to commute on the 8:30 next day.

REFERENCES

1. Lilienfeld DE, Stolley P. *Foundations of Epidemiology*. Third Edition. New York: Oxford University Press, 1994.
2. MacMahon B, Pugh TF. *Epidemiology: Principles and Methods*. Boston: Little, Brown, 1970.
3. Friedman G. *Primer of Epidemiology*. Fourth Edition. New York: McGraw-Hill, 1994.
4. Mausner JS, Kramer S. *Epidemiology: an Introductory Text*. Second Edition. Philadelphia, PA: Saunders, 1985.
5. Ahlbom A, Norell S. *Introduction to Modern Epidemiology*. Second Edition. Chestnut Hill, MA: Epidemiology Resources, 1990.
6. Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Second Edition. Boston: Little Brown, 1991.
7. Schuman SH. *Practice-Based Epidemiology*. New York: Gordon and Breach, 1986.
8. Rothman KJ, Greenland S. *Modern Epidemiology*. Second Edition. Philadelphia, PA: Lippincott-Raven, 1998.
9. Weiss N. *Clinical Epidemiology: The Study of the Outcome of Illness*. Second Edition. New York: Oxford University Press, 1996.
10. Kelsey JL, Thompson WD, Evans AS. *Methods in Observational Epidemiology*. New York: Oxford University Press, 1986.
11. Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston: Little, Brown, 1987.
12. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials*. Third Edition. Baltimore: Williams and Wilkins, 1996.
13. Hulley SB, Cummings SR. *Designing Clinical Research: An Epidemiologic Approach*. Baltimore: Williams and Wilkins, 1988.

14. Sackett DL. Bias in analytic research. *J Chron Dis* 1979; **32**: 51–63.
15. Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.
16. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; **58**: 295–300.
17. US Public Health Service. *Smoking and Health. Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington: Government Printing Office, 1964; 20.
18. Experimental carcinogenesis with tobacco smoke. In: *US Public Health Service: The Health Consequences of Smoking—Cancer. A Report of the Surgeon General*. Washington: Government Printing Office, 1982; 181.
19. Biomedical evidence for determining causality. In: *US Public Health Service: The Health Consequences of Smoking—Cancer. A Report of the Surgeon General*. Washington: Government Printing Office, 1982; 13.
20. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *New Engl J Med* 1971; **284**: 878–81.
21. Joint Commission on Prescription Drug Use. *Final Report*. Washington: Joint Commission on Prescription Drug Use, 1980.
22. Humphries TJ, Myerson RM, Gifford LM, *et al*. A unique postmarket outpatient surveillance program of cimetidine: report on Phase II and final summary. *Am J Gastroenterol* 1984; **79**: 593–6.
23. Markush RE, Seigel DG. Oral contraceptives and mortality trends from thromboembolism in the United States. *Am J Public Health* 1969; **59**: 418–34.
24. National Center for Health Statistics. *Health: United States 1982*. Hyattsville, MD: Department of Health and Human Services, 1982.
25. US Public Health Service. *Smoking and Health. Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington: Government Printing Office, 1964; A1.
26. Strom BL, Stolley PD. Vascular and cardiac risks of steroidal contraception. In: Sciarra JW, ed, *Gynecology and Obstetrics*, Vol. 6. New York: Harper and Row, 1988; Ch. 23.
27. Royal College of General Practitioners. *Oral Contraceptives and Health*. London: Pitman, 1974; Ch. 7.
28. Vessey MP, Doll R, Peto R, *et al*. A long-term follow-up study of women using different methods of contraception—an interim report. *J Biosoc Sci* 1976; **8**: 373–427.
29. Bell RL, Smith EO. Clinical trials in post-marketing surveillance of drugs. *Control Clin Trials* 1982; **3**: 61–8.
30. Mellin GW, Katzenstein M. The saga of thalidomide. *New Engl J Med* 1962; **267**: 1238–44.
31. Taussig HB. A study of the German outbreak of phocomelia. *J Am Med Assoc* 1962; **180**: 1106–14.
32. Kajii T, Kida M, Takahashi K. The effect of thalidomide intake during 113 human pregnancies. *Teratology* 1973; **8**: 163–6.

3

Sample Size Considerations for Pharmacoepidemiology Studies

BRIAN L. STROM

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

Chapter 1 pointed out that between 500 and 3000 subjects are usually exposed to a drug prior to marketing, in order to be 95% certain of detecting adverse effects that occur in between 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 pharmacoepidemiology data which are described in Part III of this book.

The approach to considering the implications of a study's sample size is somewhat different depending on whether a study is already completed or is being planned. After a study is completed, if a real finding was statistically significant, then the study had a sufficient sample size to detect it, by definition. If a finding was not statistically significant, then one can use either of two approaches. First, one can examine the resulting confidence intervals in order to

determine the smallest differences between the two study groups that the study had sufficient sample size to exclude.¹ Alternatively, one can approach the question in a manner similar to the way one would approach it if one were planning the study *de novo*. Nomograms can be used to assist a reader in interpreting negative clinical trials in this way.²

In contrast, in this chapter we will discuss in more detail how to determine a proper study sample size, from the perspective of one who is designing a study *de novo*. Specifically, we will begin by discussing how one calculates the minimum sample size necessary for a 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 one is expecting from the study.

To calculate sample sizes for a cohort study, one needs to specify five variables (see Table 3.1).^{3,4}

The first is the *alpha* (α) or *type I error* one is willing to tolerate in the study. A type I error is the probability of concluding there is a difference when in fact one 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 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 as one-tailed. If either of the study groups may be the one more likely to develop the outcome disease, and either result would be of interest, then one would specify α as two-tailed. To decide whether α should be one-tailed or two-tailed, an investigator should look at what his or her reaction would be to an answer which is statistically significant in a direction opposite to the one that would be 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 the results could come out in either direction. This is the one that is 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* one is willing to tolerate in the study. A type II error is the probability of concluding there is no difference when in fact one 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, the probability of detecting a difference if one really exists. Power is calculated as $(1 - \beta)$. Again, the more tolerant one is willing to be of type II errors, that is the higher the β , the smaller the sample size required. β is conventionally set at 0.1 or 0.2, 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 relative risk one wants to be able to detect. 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 smaller than expected, but still clinically important to the investigator. In other words, if one chooses a relative risk of 2.5, one should be comfortable with the thought that, if the relative risk is 2.4, 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 variable of interest in the unexposed control group. Again the more one asks of a study, the larger the sample size

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 relative risk to be detected
(4) Incidence of the disease in the unexposed control group	(4) Prevalence of the exposure in the undiseased control group
(5) Ratio of unexposed controls to exposed study subjects	(5) Ratio of undiseased controls to diseased study subjects

one needs. The rarer the outcome of interest, the larger the sample size needed.

The fifth variable one needs to specify is the number of unexposed control subjects to be included in the study for each exposed study subject. A study has the most statistical power for a given number of study subjects if it has the same number of controls as exposed subjects. However, sometimes the number of exposed subjects is limited and, therefore, inadequate to provide sufficient power to detect a relative risk of interest. In that case, additional power can be gained by increasing the number of controls alone. Doubling the number of controls, that is including two controls for each exposed subject, results in a modest increase in the statistical power, but it does not double it. Including three controls for each exposed subject increases the power further. However, the increment in power due to increasing the ratio of control subjects to exposed subjects from 2:1 to 3:1 is smaller than the increment in power due to 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, if one were to 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 give one a power of 0.80, or an 80% chance of detecting a difference of that size. 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 control groups should not be aggregated in the analysis. 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 which is probably the most often used is modified from Schlesselman:³

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

where p is the incidence of the disease in the unexposed, R is the minimum relative risk to be detected, α is the type I error rate that is acceptable, β is the type II error rate that 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 calculated using this formula is presented in Appendix A. In Tables A1–A4 we have assumed an α (two-tailed) of 0.05, a β of 0.1, and control to exposed ratios of 1:1, 2:1, 3:1, and 4:1, respectively. Tables A5–A8 are similar, except they assume a β of 0.2. Each table presents the number of exposed subjects needed to detect any of a number of specified relative risks, for

outcome variables that occur with a number of specified incidence rates.

For example, what if one wanted to investigate a new nonsteroidal anti-inflammatory drug that was 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 which occurs, say, in 0.1% of all unexposed individuals? If one wanted to design a study with one control per exposed subject to detect a relative risk of 2.0 for this outcome variable, assuming an α (two tailed) of 0.05 and a β of 0.1, one could look in Table A1 and see that it would require 31 483 exposed

subjects, as well as an equal number of unexposed controls. If one were less concerned with missing a real finding, even if it were there, one could change β to 0.2, and the required sample size would drop to 23 518 (see Table 3.2 and Table A5). If one wanted to minimize the number of exposed subjects needed for the study, one could include up to four controls for each exposed subject (Table 3.2 and Table A8). This would result in a sample size of 13 402, with four times as many controls, a total of 67 010 subjects. Finally, if one considers it inconceivable that this new drug could *protect* against liver disease and is not interested in that outcome, then one might use a one-tailed α , resulting in a somewhat lower sample size: 10 728, again with four times as many controls. Much smaller sample sizes are needed to detect relative risks of 4.0 or greater, and these are also presented in Table 3.2.

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	3 104	3 104
	0.01	0.05 (2-tailed)	0.2	2	1	2 319	2 319
	0.01	0.05 (2-tailed)	0.2	2	4	1 323	5 292
	0.01	0.05 (1-tailed)	0.2	2	4	1 059	4 236
	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	5 823	5 823
	0.001	0.05 (2-tailed)	0.2	4	1	4 350	4 350
	0.001	0.05 (2-tailed)	0.2	4	4	2 253	9 012
	0.001	0.05 (1-tailed)	0.2	4	4	1 829	7 316
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

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 outcome variable, only 3 104 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 relative risk (odds ratio) one wants to be able detect. These are discussed in the section on cohort studies, above.

In addition, in a case-control study one selects subjects on the basis of 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 on the basis of 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, for cohort studies we are concerned about the ratio of the number of unexposed control subjects to the number of exposed study subjects. For a case-control study, we are

analogously interested in 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 give one 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 reference):³

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

where R , α , β , $Z_{1-\alpha}$, and $Z_{1-\beta}$ are as above, p is the prevalence of the exposure in the control group, 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 the Appendix. In Tables A9-A12 we have assumed

an α (two-tailed) of 0.05, a β of 0.1, and control to case ratios of 1:1, 2:1, 3:1, and 4:1, respectively. Tables A13–A16 are similar, except they assume a β of 0.2. 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 was about to be marketed but pre-marketing 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 exposed 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 would be used nearly as commonly as ibuprofen, by perhaps 1% of the population, then one could look in Table A9 and see that it would require 3210 diseased subjects, as well as an equal number of undiseased controls. If one were less concerned with missing a real association, even if it existed, one could change β to 0.2, and the required sample size would drop to 2398 (see Table 3.3 and Table A13). If one wanted to minimize the number of diseased subjects needed for the study, one could include up to four controls for each exposed subject (Table 3.3 and Table A16). This would result in a sample size of 1370, with four times as many controls. Finally, if one considered it inconceivable that this new drug could *protect* against liver disease, then one might use a one-tailed α , resulting in a somewhat lower sample size: 1 096, again with four times as many

Table 3.3 Examples of sample sizes needed for a case control study

Hypothetical drug	Prevalance rate assumed in undiseased	α	β	Odds ratio to be detected	Control: case ratio	Sample size needed in case group	Sample size control group
Ibuprofen	0.01	0.05 (2-tailed)	0.1	2	1	3 210	3 210
	0.01	0.05 (2-tailed)	0.2	2	1	2 398	2 398
	0.01	0.05 (2-tailed)	0.2	2	4	1 370	5 480
	0.01	0.05 (1-tailed)	0.2	2	4	1 096	4 384
	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	5 856	5 856
	0.001	0.05 (2-tailed)	0.2	4	1	4 375	4 375
	0.001	0.05 (2-tailed)	0.2	4	4	2 266	9 064
	0.001	0.05 (1-tailed)	0.2	4	4	1 840	7 360
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

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, 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 exposed 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 which 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 of prazosin, the investigators collected a case series of 10 000 newly exposed subjects recruited through the manufacturer's sales force, to quantitate better the incidence of first dose syncope, which was a well recognized adverse effect of this drug.^{5,6} Case series are 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 chemically closely related to metiamide, another H-2 blocker, which had been removed from the market in Europe because it caused agranulocytosis. This study also collected 10 000 subjects. It found only two cases of neutropenia, one in a patient also receiving chemotherapy. There were no cases of agranulocytosis.⁷

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

Table A17 in Appendix A presents a table useful for making this calculation.⁸ In order to apply this table, one first calculates the incidence rate observed from the study's results, that is the number of subjects who develop the disease of interest during the specified time interval, divided by the total number of individuals in the population at risk. For example, if three cases of liver disease were observed in a population of 1000 patients exposed to a new nonsteroidal anti-inflammatory drug during a specified period of time, the incidence would be 0.003. The number of subjects who develop the disease is the "Observed number on which estimate is based (*n*)" in Table A17. In this example, it is 3. The lower boundary of the 95% confidence interval for the incidence rate is then the corresponding "Lower limit factor (L)" multiplied by the observed incidence rate. In 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"⁴ multiplied by the observed incidence rate. In the above example, this would be $2.92 \times 0.003 = 0.00876$. In other words, the incidence rate (95% confidence interval) would be 0.003 (0.000618–0.00876). Thus, the best estimate of the incidence rate would

be 30 per 10 000, but there is a 95% chance that it lies between 6.18 per 10 000 and 87.6 per 10 000.

In addition, a helpful simple guide is the so called "rule of threes," useful in the common situation where no events of a particular kind are observed.⁸ Specifically, if no events of a particular type 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 occurs less often than three 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 occurs in no more than three 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 occur less often than three in 10 000 exposed individuals, or that they have an incidence rate of less than 0.0003. In other words, events not detected in the study occur less often than one in 3333 subjects.

DISCUSSION

The discussions about sample size determinations in cohort and case-control studies above assume one is able to obtain information on each of five variables. 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 whole 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 one will study will be similar. If this is not an appropriate assumption, or if no such data exist in the medical literature, one is left with two alternatives. The first, and better, alternative is to conduct a small pilot study within one's population, in order to measure the information one needs. The second is simply to guess. In the second case, one should consider what a reasonable higher guess and a reasonable lower guess might be, as well, to see if one's sample size should be increased to take into account the imprecision of one's estimate.

Finally, what if one is studying multiple outcome variables, each of which differs in the frequency one expects in the control group? In that situation an investigator might base the study's sample size on the variable that leads to the largest requirement, and note that the study will have even more power for the other outcome variables. It is usually better to have a somewhat larger expected sample size than the minimum, in any case, to allow some leeway if any of the underlying assumptions are wrong. This also will permit subgroup analyses with adequate power. In fact, if there are important subgroup analyses that represent *a priori* hypotheses 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 examine his or her final assumptions closely, asking whether, given the compromises made, the study is still worth mounting.

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

All of the previous discussions have focused on calculating a minimum necessary sample size. This is the usual concern. However, two other issues specific to pharmacoepidemiology are important to consider as well. First, one of the main advantages of postmarketing 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, of course, need to be very much larger. Case-control studies can be much smaller, but generally need to recruit cases and controls from a source population of equivalent size. These are not totally arbitrary figures, but are based on the principles described above, applied to the questions that 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, a number of new techniques have been

developed, which are described in Part III 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. Also, 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.

REFERENCES

1. Makuch RW, Johnson MF. Some issues in the design and interpretation of "negative" clinical trials. *Arch Intern Med* 1986; **146**: 986-9.
2. Young MJ, Bresnitz EA, Strom BL. Sample size nomograms for interpreting negative clinical studies. *Ann Intern Med* 1983; **99**: 248-51.
3. Schlesselman JJ. Sample size requirements in cohort and case-control studies of disease. *Am J Epidemiol* 1974; **99**: 381-4.
4. Stolley PD, Strom BL. Sample size calculations for clinical pharmacology studies. *Clin Pharmacol Ther* 1986; **39**: 489-90.
5. Graham RM, Thornell IR, Gain JM, Bagnoli C, Oates HF, Stokes GS. Prazosin: the first dose phenomenon. *Br Med J* 1976; **2**: 1293-4.
6. Joint Commission on Prescription Drug Use. *Final Report*. Washington: Joint Commission on Prescription Drug Use, 1980.
7. Gifford LM, Aeugle ME, Myerson RM, Tannenbaum PJ. Cimetidine postmarket outpatient surveillance program. *J Am Med Assoc* 1980; **243**: 1532-5.
8. Haenszel W, Loveland DB, Sirken MG. Lung cancer mortality as related to residence and smoking history. I. White males. *J Natl Cancer Inst* 1962; **28**: 947-1001.
9. Cohen J. *Statistical Power Analysis for the Social Sciences*. New York: Academic Press, 1977.

4

Basic Principles of Clinical Pharmacology Relevant to Pharmacoepidemiology Studies

PATRICIA McGETTIGAN

Faculty of Medicine and Health Sciences, The University of Newcastle, Newcastle, NSW, Australia

DAVID A. HENRY

Faculty of Medicine and Health Sciences, The University of Newcastle, Newcastle, NSW, Australia

SEAN HENNESSY

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

INTRODUCTION

Clinical pharmacology is defined classically as the study of the effects of drugs in humans. More broadly, it also considers 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 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, hence 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. Of course, neither approach would be justified if responses to drugs were totally predictable. Pharmacoepidemiology is the application of epidemiologic methods to the subject matter of clinical pharmacology. 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, or they may be due to a number of 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 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. In addition, 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 34).

In 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 disciplines of clinical pharmacology and pharmacotherapeutics have 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 that accomplish this. The emphasis of this chapter will be on concepts that are likely to be important in conducting and understanding pharmacoepidemiology 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 in determining variability, the role of pharmacokinetics in determining variability, and the importance of the human factor 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 alfa-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, and preservatives) that are chosen, among other concerns, because they are 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.¹

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 name. This is potentially of great concern to any pharmacoepidemiologist interested in studying the effects of non-prescription drugs.

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. 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.²

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 (beta) 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.³ 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 parenterally.⁴ 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 be delivered to the patient is rarely at issue; if need be, the dose can simply increased. Milligram potency is rarely an important consideration in therapeutic drug use.

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.

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. 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 non-experimental. Further, there is considerable information about determinants of patients' responses to drugs in general. In this section, we present the effects of adaptive responses, age, disease states, and concomitant drugs in determining variability in drug response.

EFFECTS OF ADAPTIVE RESPONSES

It is a general rule of pharmacology that pharmacodynamic responses tend to be followed by adaptive responses which, crudely put, are the body's attempt to "overcome" 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 blocking⁵ agents, such as propranolol, resulting in increased numbers of active β -receptors. If the beta-blocking drug is withdrawn rapidly, a large number of β -receptors become available to bind to their natural ligands, norepinephrine and epinephrine. This can produce tachycardia, hypertension, and worsening angina—the so-called " β -blocker withdrawal syndrome".⁵

In some cases, the mechanisms of apparent adaptive responses have not yet been fully explored. For example, among subjects taking NSAIDs, endoscopic studies have documented gastrointestinal mucosal damage within days of commencing treatment.⁶² Endoscopic investigation of patients chronically exposed to aspirin found that the mucosal damage appeared to resolve over time.⁷ While this suggested continued exposure promoted gastric adaptation, the mechanism by which this might occur was unclear. Although endoscopic evidence of mucosal damage represents a surrogate for the outcome of clinical interest, gastric ulceration and its complications of bleeding, perforation, and stenosis, epidemiological studies 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.^{63,64} More recent evidence suggests the existence of an adaptive mucosal response is questionable. In a record linkage study, MacDonald *et al.* 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.⁶⁵ These findings were at variance with the earlier studies, but as the prospective design of the study by MacDonald *et al.* is better suited to the examination of temporal effects, their results are likely to be valid.

NSAID-induced gastric mucosal damage appears to be related to the inhibition of gastric prostaglandin synthesis by the enzyme cyclo-oxygenase. Prostaglandins are a large family of 'local hormones' or 'autocoids' (Greek, *autos*: *self*, *akos*: *medicinal agent or remedy*) with diverse haematological, smooth muscular, gastric, cardiac, renal, nervous system, endocrine, and metabolic effects. COX is known to exist in at least two isoforms. COX-1, physiological COX, is believed to predominate in the stomach, generating the prostaglandins that protect the gastric mucosa from acid damage. COX-2, inducible COX, is associated with the production of inflammatory prostaglandins. NSAIDs in common use, such as ibuprofen, diclofenac, naproxen, are non-selective COX inhibitors, blocking formation of both COX-1 and COX-2. It has been postulated that the anti-inflammatory actions of these agents are due to their inhibitory effects on COX-2 formation, their toxicity, in particular that relating to the GI tract, to COX-1 inhibition. The distinction between the COX isoforms led to a search for drugs that spare COX-1 (COX-2 inhibitors) in the expectation that such agents would be free of the serious GI toxicity caused by non-selective NSAIDs. The first COX-2 inhibitors, celecoxib and rofecoxib, have now been marketed and others are being developed. Endoscopic studies found that both celecoxib and rofecoxib were associated with significantly less mucosal damage than non-selective agents.⁶⁶⁻⁶⁸ The rates of clinically significant complications including symptomatic ulceration, perforation, bleeding, or stenosis were low, and thereby limited the trials' capacity to quantify any advantage of COX-2 inhibitors in relation to these outcomes. Data from large numbers of patients will be required and this is most likely to be provided from post-marketing surveillance and epidemiological studies.

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 is somewhat counterintuitive, the elderly are often *equally* or even *less* sensitive to the primary pharmacologic effects of some drugs than are the young. 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.¹⁰ The extent to which this is due to elevated levels of plasma catecholamines is not clear.
- Elegant experimental work has demonstrated that elderly subjects have a blunted primary electrophysiologic response to the calcium channel blocking drug verapamil.¹¹
- 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.¹⁰
- 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.¹⁰

How may the last two observations be reconciled? The probable 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.¹⁰ 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 which 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 hyponatraemia or volume overload.

Postural hypotension (the sudden drop in blood pressure that occurs with standing or sitting up) is frequently symptomatic in the elderly and 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, anti-hypertensives, diuretics, and levodopa.

EFFECTS OF DISEASE STATES

The effects of disease states on pharmacodynamics have not been widely studied. 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.^{12,13} 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 occur 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.¹⁴

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.¹⁵ 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 (for example, 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 most clinically 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. 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 methyldopa. A "serotonergic syndrome" (consisting of mental changes, muscle rigidity, hypertension, tremor, and diarrhea) may be induced in some patients given combinations of proserotonergic drugs such as

selective serotonin reuptake inhibitors (SSRIs), 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 nonselective β -antagonists (bronchoconstrictors).

Sildenafil (Viagra®) is a selective inhibitor of 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 natural erectile response to sexual stimulation in men with erectile dysfunction.¹⁶ 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. Although concomitant use is contraindicated by the manufacturer, a number of sildenafil-associated deaths are considered 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.¹⁷ The use of amphetamine-like appetite suppressants, mainly fenfluramine and dexfenfluramine, has also been associated with primary pulmonary hypertension.¹⁸ It has been postulated that these unintended effects were due to serotonin accumulation as a consequence of both increased release and reduced removal. 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 have produced 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 pharmacoepidemiology 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 above, pharmacokinetics is the science that describes the time course of the absorption, distribution, and elimination of drugs within the body. 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

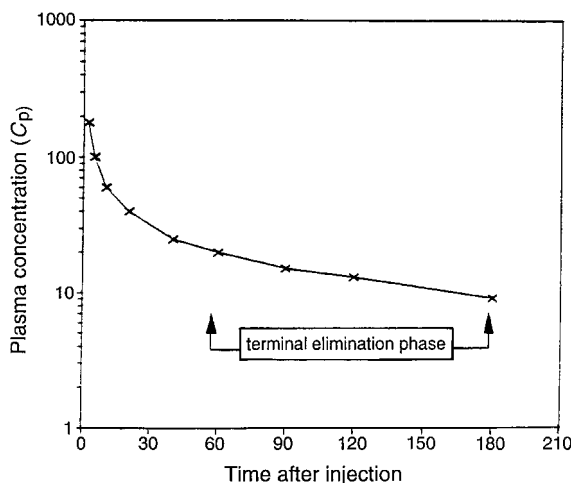


Figure 4.1. Plasma concentration/time curve of a hypothetical drug after intravenous administration. Note the early rapid decline in blood levels which reflects distribution of the drug, a consequence of its lipid solubility and degree of protein binding. The terminal phase of the concentration/time 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 min.

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 semilog scale. It was observed that, for some drugs, the initial portion of the log concentration/time 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 into which drug molecules are said to distribute, and is represented by a given linear component of the log concentration/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.

The *central compartment* is the body fluid in which the drug concentration is being measured (plasma, in the case of Figure 4.1). The initial, rapid decline of measured drug concentration is attributed to distribution of drug molecules from

the central compartment into one or more *peripheral compartments*, and is 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 litres, and is the volume of fluid 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 litres. This is because the drug partitions readily into fatty tissue, leaving little measurable drug in the blood stream.

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 hours.¹ Because of the linearity of the terminal elimination phase, the time that it takes for any given drug concentration to decline to half of this 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 of the body fluid 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 litres per hour. Clearance can be calculated by taking the product of the apparent volume of distribution and the elimination rate constant.

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 is equal to 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 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 the longer half-life results in a longer time to achieve steady state concentrations and a tendency to accumulate.

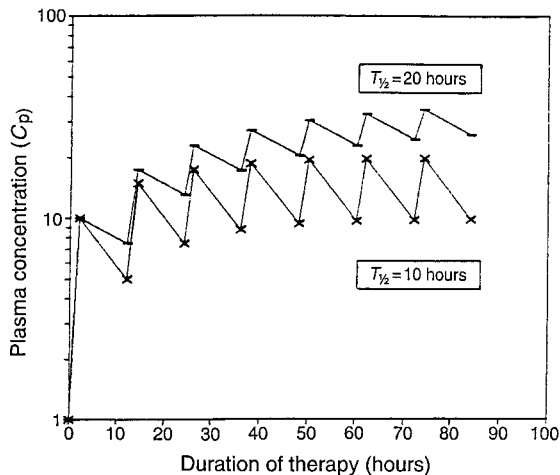


Figure 4.2. Profile of plasma drug concentrations for two hypothetical drugs during repeated 12 hourly dosing schedule. The lower curve relates to a drug with a half-life of 10 hours; steady state concentrations are achieved after 50–60 hours. In contrast, the upper curve relates to a drug with a half-life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

EFFECTS OF VARIATION IN DRUG ABSORPTION

In clinical practice, most drugs are administered by mouth. Because they are small and at least partially lipid soluble, most drug molecules 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 physico-chemical 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 can affect the extent and rate of gastrointestinal absorption of drugs and, if they have an effect, both will tend to *decrease* absorption. As a result, altered gastrointestinal absorption is unlikely to be a cause of increased susceptibility to adverse drug effects, and has traditionally been of little concern in pharmacoepidemiology research. An important exception exists where different formulations of the same drug have dramatically different absorption profiles. 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 severe hypotension, presumably due to rapid absorption leading to elevated peak concentrations of the drug, with subsequent marked vasodilatation.¹⁹

Also of interest to pharmacoepidemiologists is the systemic absorption of drugs that may occur following unintended absorption via other routes, such as transdermally, following administration by metered dose inhaler, or ocular instillation. 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. Unfortunately, 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,²⁰ and in another instance,

following the inadvertent contamination of talcum powder with the anticoagulant warfarin.²¹ 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.²² Likewise, β -blocking drugs instilled into the eyes can travel down the nasolachrymal duct to be swallowed and absorbed, inducing bronchospasm or exacerbation of congestive heart failure in susceptible individuals.²³

In summary, because variability in the absorption of oral dosage forms of drugs from the gastrointestinal tract typically *reduces* absorption, it is rarely important in pharmacoepidemiology studies of 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 into the central compartment, 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 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 higher concentration of drug than tissues that are poorly perfused.

Protein binding is an aspect of drug distribution that receives considerable attention. 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.²⁴ 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.

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 pharmacoepidemiology findings. However, alterations in drug distribution rarely produce clinically important effects. Pharmacokinetic explanations for variability in response usually involve a change in clearance of drugs from the body.

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, 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, highly 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, 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 overdose.

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).

Blood 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 blood 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

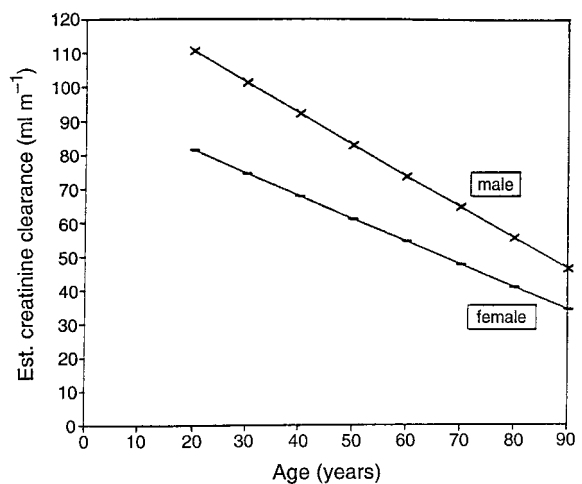


Figure 4.3. Change in estimated creatinine clearance (Cockcroft and Gault formula²⁵) with age in a male and a female who maintain a serum creatinine of 0.1 mmol l^{-1} (NR $0.07 - 0.12 \text{ mmol l}^{-1}$), throughout their lives. In estimating creatinine clearance, it was assumed that the male maintained 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.

pharmacoepidemiology studies of drugs for which this is the principal route of elimination from the body. Some linked pharmacoepidemiology databases include laboratory files, enabling estimates of renal function to be made and included in analyses of outcomes. Consequently, it is important to recognize that blood creatinine concentrations must be adjusted for age and body 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.²⁵

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.²⁶ Accumulation of these agents can produce serious adverse effects.

In cases where filtration pressure is maintained by vasoconstriction of the post-glomerular efferent arteriole, angiotensin converting enzyme inhibitors (ACEIs) 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, cyclosporin, induces vasoconstriction of the afferent glomerular arteriole in a dose related and reversible fashion. An increased risk of acute renal failure exists when cyclosporin 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 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 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; pronounced “sip”) 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

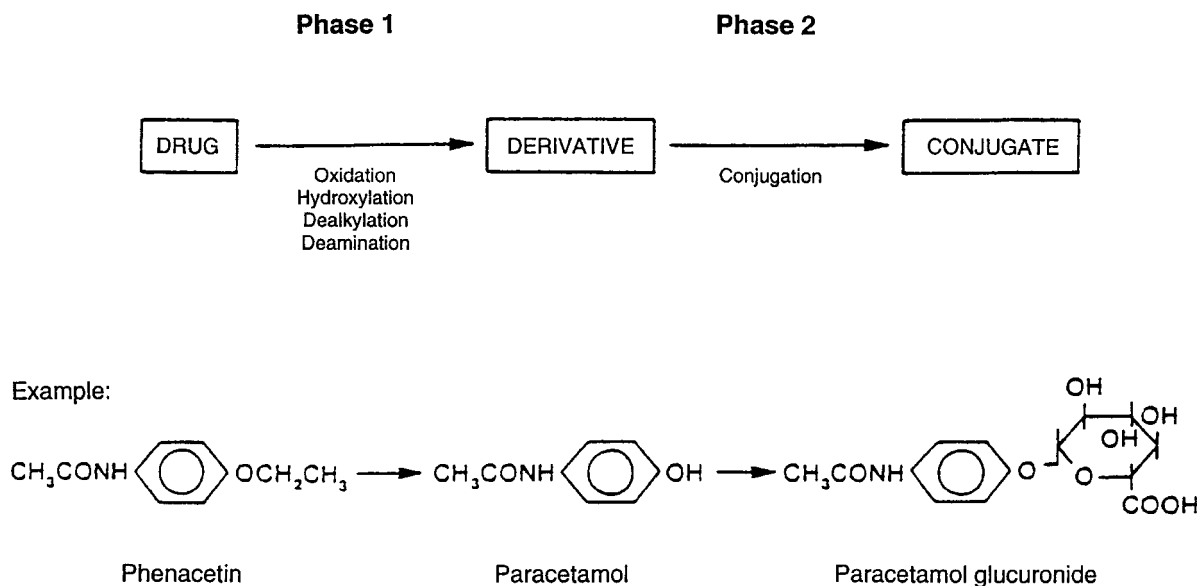


Figure 4.4 (Figure and legend reproduced with the permission of HP Rang and MM Dale³ and the publishers Churchill Livingstone, Edinburgh.) Phase I and II reactions often occur sequentially. Phase I reactions usually consist of oxidation, reduction, hydrolysis, and products are often more reactive, and 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.

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,²⁶ 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,²⁷ while temazepam metabolism is relatively unaffected by these factors.²⁸

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 anticoagulants 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 five- to ten-fold variations.²⁹

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.²⁹ 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.³⁰

Acetylation polymorphism affects the metabolism of a number of drugs in addition to isoniazid; these include some sulfonamides (including sulfasalazine), hydralazine, procainamide, dapsone, nitrzepam, 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. This is particularly the case for drugs that have a narrow therapeutic index (that is, little difference between toxic and therapeutic concentrations).

Hydroxylation polymorphism was identified in 1977.³¹ 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), anti-arrhythmics (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.

CYP2D19 polymorphism is described in 3% of Caucasians and 15–25% of Asians who have a deficient capacity to hydroxylate mephenytoin.³¹ CYP2D19 also catalyzes the metabolism of commonly used drugs such as diazepam, omeprazole, propranolol, some tricyclic antidepressants, and moclobemide.

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 interindividual and interethnic 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 compared debrisoquine hydroxylation in a group of subjects who had developed lactic acidosis while taking phenformin with the expected distribution in the Swedish population.³² This study illustrates the potential for investigating groups of individuals who display apparently idiosyncratic reactions to certain drugs. More recent 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.^{33,34}

The large series of well validated case reports held by many spontaneous reporting systems represent fertile areas for this type of research (see Chapters 10 and 11). The use of biochemical tests in concert with epidemiologic methods in order to predict variability in drug response is a promising new area of research.

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 tend to be spared. Other individuals, such as those with biliary obstruction or acute viral hepatitis, may have remarkably 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.³⁵

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 and this should be considered as an explanation for unexpected pharmacoepidemiology findings. For example, several metabolites of carbamazepine contribute to its pharmacologic activity.³⁶ The hydroxyl metabolite of propranolol has similar activity to its parent compound.³⁷ 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.³⁷ Likewise, the acetyl derivative of the antiarrhythmic 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-acetylbenzoquinoneimine 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 are thought to be more prone to develop liver damage. Acute doses of alcohol, on the other hand, markedly inhibit the formation of the metabolite and may actually be protective.

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.

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 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. 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 ethynylestradiol containing oral contraceptives can result from the CYP450 enzyme inducing effects of some anti-epileptic medications.^{26,39} 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.³ It can take several weeks to reach its peak, 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.²⁵

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.⁴⁰ In contrast, omeprazole has been shown to inhibit the metabolism of diazepam and phenytoin, but not of propranolol.⁴¹⁻⁴³ Erythromycin is a clinically important enzyme inhibitor, well known for its effects on theophylline metabolism. More recently, erythromycin and other macrolide antibiotics have been shown to inhibit the metabolism of the antihistamines terfenadine and astemizole, with consequent adverse effects on cardiac conduction.⁴⁴ The calcium antagonists diltiazem and verapamil (but not nifedipine) increase cyclosporin plasma concentrations, but with relative sparing of nephrotoxicity, and the interaction has been used in clinical practice to produce an immunosuppressive concentration of cyclosporin at a lower ingested dose.⁴⁵ Drug cost savings of 14-48%, attributable to the use of calcium antagonists, have been reported in transplant pharmacotherapy. The novel calcium antagonist, mibefradil, was withdrawn from the market in 1998 as a consequence of its interactions with concomitantly administered β -blockers and dihydropyridine calcium antagonists which resulted in bradycardia, infarction, cardiogenic shock, and death.⁴⁶

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.⁴⁷ A similar effect of biflavonoids on cyclosporin concentrations has been observed and has been utilised to reduce the doses, and therefore the side effects and costs, of cyclosporin therapy. Both calcium antagonists and cyclosporin are metabolized by the CYP450 isoenzyme CYP3A4 which is present in the gut wall and in the liver, and the biflavonoids inhibit its activity. 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. In such cases, a small increase in plasma concentration may have a disproportionate clinical effect. 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. Cyclosporin 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 not produce significant adverse effects.

THE IMPORTANCE OF THE HUMAN FACTOR: PRESCRIBER AND CONSUMER BEHAVIOR

Human behavior may be a greater source of variability in patterns of drug use than any other factor considered so far in this chapter. In conducting pharmacoepidemiology studies, it is important to give cognisance to its impact upon

observed prescribing 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 advertising.

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 diastolic 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 be therefore 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 34, 43, and 44).

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, such as 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 inappropriate first choice 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.^{48,49} 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.⁴⁹ 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.⁴⁸

RISKS AND BENEFITS OF TREATMENT

The risks 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.⁵⁰ Perception of risk 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.⁵¹

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.^{52,53} Promotional

materials from pharmaceutical companies frequently present the benefits of treatments in relative as opposed to absolute terms. As the former are numerically higher in most cases than the latter, they 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.

ECONOMIC INFLUENCES

Economic influences, exerted from various sources, may influence drug use and therefore the interpretation of pharmacoepidemiology studies. 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.⁵⁴

In England in the past decade, the Department of Health has introduced several schemes intended to contain the costs of National Health Service prescribing. These have 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. For example, practices tended to exceed their indicative budgets,⁵⁵ but fundholders appeared to achieve more effective cost containment than non fundholders, mainly through eliminating unnecessary prescribing, moving to generic prescribing, and altering drug choice within therapeutic classes.⁵⁶

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 (see also Chapters 27 and 35).

Other approaches intended to contain prescribing costs have included national formularies and limited lists, patient co-payment, generic substitution, 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.

Prescribers themselves may have a pecuniary interest in prescribing. Fee-per-item methods of physician remuneration have been found to encourage a higher use of services.^{57,58} 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.^{59,60} 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 most 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.

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 (for example, nurses at asthma or diabetic clinics), sponsorship to attend international speciality 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 nonprescription 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 nonprescription NSAID.

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 postmenopausal estrogen therapy, consumers were shown to make greater use of other medical services and to have higher levels of exercise than nonconsumers.⁶¹ 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 34).

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 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.

REFERENCES

1. American Academy of Pediatrics Committee on Fetus and Newborn Committee on Drugs. Benzyl alcohol: toxic agent in neonatal units. *Pediatrics* 1983; **72**: 356.
2. Nayler WG. *Calcium Antagonists*. London: Academic Press, 1988.
3. Rang HP, Dale MM. *Pharmacology*, 2nd edn. Edinburgh: Churchill Livingstone, 1991.
4. Jaffe JH, Martin WR. Opioid analgesics and antagonists. In: Gilman AG, Rall TW, Nies AS, Taylor P, eds, *The Pharmacological Basis of Therapeutics*, 8th edn. New York: Pergamon Press, 1990; 497.
5. Kendall MJ, Beeley L. Beta-adrenoceptor blocking drugs: adverse reactions and drug interactions. *Pharmacol Ther* 1983; **21**: 351–69.
6. Dequeker J, Hawkey C, Kahan A, *et al.*, on behalf of the SELECT Study Group. Improvement in gastrointestinal tolerability of the selective cyclooxygenase (COX)-2 inhibitor, meloxicam, compared with piroxicam: results of the safety and efficacy large-scale evaluation of COX-inhibiting therapies (SELECT) trial in osteoarthritis. *Br J Rheumatol* 1998; **37**: 946–51.
7. Graham DY, Smith JL, Spjut HJ, Torres E. Gastric adaptation. *Gastroenterology* 1988; **95**: 327–33.
8. Griffin MR, Piper JM, Daugherty JR, *et al.* Nonsteroidal anti-inflammatory drug use and increased

- risk for peptic ulcer disease in elderly persons. *Ann Intern Med* 1991; **114**: 257–63.
9. Carson JL. A case study: non-steroidal anti-inflammatory drugs and gastrointestinal bleeding. *J Rheumatol* 1988; **17** (Suppl.): 24–7.
 10. Vestal RE, Wood AJ, Shand DG. Reduced beta adrenoreceptor sensitivity in the elderly. *Clin Pharmacol Ther* 1979; **26**: 181–6.
 11. Abernethy DR, Schwartz JB, Todd EL, *et al.* Verapamil pharmacodynamics and disposition in young and elderly hypertensive patients. Altered electrocardiographic and hypotensive responses. *Ann Intern Med* 1986; **105**: 329–36.
 12. Podoisky DK, Isselbacher KJ. Derangements of hepatic metabolism. In: Wilson JD, Braunwald E, Isselbacher KJ, *et al.*, eds, *Harrison's Principles of Internal Medicine*, 12th edn. New York: McGraw-Hill, 1991; 1311–7.
 13. Anon. Morphine. In: Dollery C, ed., *Therapeutic Drugs*, Vol. 2. Edinburgh: Churchill Livingstone, 1991; M225–33.
 14. Grimm G, Ferenci P, Katzenschlager R, *et al.* Improvement of hepatic encephalopathy treated with flumazenil. *Lancet* 1988; **2**: 1392–4.
 15. Clive DM, Stoff JS. Renal syndromes associated with non-steroidal anti-inflammatory drugs. *New Engl J Med* 1984; **310**: 563–72.
 16. Goldstein I, Lue TF, Paduea-Nathan H, *et al.* Oral sildenafil in the treatment of erectile dysfunction. *New Engl J Med* 1998; **338**: 1397–404.
 17. Khan MA, Herzog CA, St Peter JV, *et al.* The prevalence of cardiac valvular insufficiency assessed by transthoracic echocardiography in obese patients treated with appetite suppressant drugs. *New Engl J Med* 1998; **339**: 713–8.
 18. Abenheim L, Moride Y, Brenot F, *et al.*, for the International Primary Pulmonary Hypertension Group. Appetite-suppressant drugs and the risk of primary pulmonary hypertension. *New Engl J Med* 1996; **335**: 609–16.
 19. Anon. Adalat tablets. Approved product information. In: Thomas J, ed., *Prescription Products Guide 1992*. Hawthorne Publishing Company: Australian Pharmaceutical, 1992; 237–8.
 20. Larregue M, Bressieux JM, Laidet B, Titi A. Caustic diaper dermatitis and toxic encephalopathy following the application of talc contaminated by hexachlorophene. *Ann Pediatr (Paris)* 1986; **33**: 587–92.
 21. Martin-Bouyer G, Khanh NB, Linh PD, *et al.* Epidemic of haemorrhagic disease in Vietnamese infants caused by warfarin-contaminated talcs. *Lancet* 1983; **1**: 230–2.
 22. Law CM, Preece MA, Warner JO. Nocturnal adrenal suppression in children inhaling beclomethasone dipropionate (letter). *Lancet* 1987; **1**: 1321.
 23. Anon. Timoptol. Approved product information. In: Thomas J, ed., *Prescription Products Guide*. Hawthorne: Australian Pharmaceutical, 1992; 17401.
 24. Birkett D. Drug protein binding. *Aust Prescr* 1992; **15**: 56–7.
 25. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron* 1976; **16**: 31–41.
 26. Stockley IH. *Drug Interactions*, 2nd edn. Oxford: Blackwell Scientific Publications, 1991.
 27. Klotz U, Avant GR, Hoyumpa A, *et al.* The effect of age and liver disease on the disposition and elimination of diazepam in adult man. *J Clin Invest* 1975; **55**: 347–59.
 28. Ghabrial H, Desmond PV, Watson KJ, *et al.* The effects of age and chronic liver disease on the elimination of temazepam. *Eur J Clin Pharmacol* 1986; **30**: 93–7.
 29. Graham-Smith DG, Aronson JK. Pharmacokinetics. In: *Oxford Textbook of Clinical Pharmacology and Drug Therapy*, 2nd edn. Oxford: Oxford University Press, 1992; 94–103.
 30. McEvoy GK, ed. *AHFS Drug Information*. Bethesda, MD: American Society of Hospital Pharmacists, 1993; 355.
 31. Mahgoub A, Idel JR, Dring LG, *et al.* Polymorphic hydroxylation of debrisoquine in man. *Lancet* 1977; **2**: 584.
 32. Wiholm B-E, Alvan G, Bertilsson L, *et al.* Hydroxylation of debrisoquine in patients with lactic acidosis after phenforinin. *Lancet* 1981; **1**: 1098–9.
 33. Farrell G, Prendergast D, Murray M. Halothane hepatitis. Detection of a constitutional susceptibility factor. *New Engl J Med* 1985; **313**: 1310–4.
 34. Gennis MA, Vemuri R, Burns EA, *et al.* Familial occurrence of hypersensitivity to phenytoin. *Am J Med* 1991; **91**: 631–4.
 35. Jusko WJ, Eaten ML. Factors affecting theophylline disposition. In: McLeod SM, Isles A, eds, *Theophylline Therapy Update*. Mississauga: Astra Pharmaceuticals, 1982; 19–29.
 36. Anon. Carbamazepine. In: Dollery C, ed., *Therapeutic Drugs*, Vol. 1. Edinburgh: Churchill Livingstone, 1991; 49–53.
 37. Anon. Propranolol. In: Dollery C, ed., *Therapeutic Drugs*, Vol. 1. Edinburgh: Churchill Livingstone, 1991; 272–8.
 38. Osborne RJ, Joel SP, Slevin ML. Morphine intoxication in renal failure: the role of morphine-6-glucuronide. *Br Med J* 1986; **292**: 1548–9.
 39. Janz D, Schmidt D. Anti-epileptic drugs and failure of oral contraceptives (letter). *Lancet* 1974; **1**: 1113.
 40. Anon. Gastrointestinal drugs. In: McEvoy GK, ed., *AHFS Drug Information*. Bethesda, MD: American Society of Hospital Pharmacists, 1990; 1666–70.
 41. Gugler R, Jensen JC. Omeprazole inhibits elimination of diazepam (letter). *Lancet* 1984; **1**: 969.

42. Prichard PJ, Walt RP, Kitchingman GK, *et al.* Oral phenytoin kinetics during omeprazole therapy. *Br J Clin Pharmacol* 1987; **24**: 534–5.
43. Henry D, Brent P, Whyte I, *et al.* Propranolol steady state pharmacokinetics are unaltered by omeprazole. *Eur J Clin Pharmacol* 1987; **33**: 369–73.
44. Anon. Ventricular arrhythmias due to terfenadine and astemizole. Current problems. *Committee on Safety of Medicines* 1992; **35**: 1–2.
45. McCombs CC, deShazo RD. Immune system disorders. In: Speight TM, Holford NHG, eds, *Avery's Drug Treatment*, 4th edn. Auckland: Adis International, 1997; 1308–9.
46. Mullins ME, Horowitz BZ, Linden DHJ, *et al.* Life-threatening interactions of mibefradil and beta-blockers with dihydropyridine calcium channel blockers. *J Am Med Assoc* 1998; **280**: 157–8.
47. Bailey DG, Arnold JMO, Spence JD. Grapefruit juice and drugs: how significant is the interaction? *Clin Pharmacokinet* 1994; **26**: 91.
48. Britten N, Ukoumunne O. The influence of patients' hopes of receiving a prescription on doctors' perceptions and the decision to prescribe: a questionnaire survey. *Br Med J* 1997; **315**: 1506–10.
49. Cockburn J, Pit S. Prescribing behaviour in clinical practice: patients' expectations and doctors' perceptions of patients' expectations—a questionnaire study. *Br Med J* 1997; **315**: 520–3.
50. Nikolajevic-Sarunac J, Henry DA, O'Connell DL, Robertson J. Effects of information framing on general practitioners' intentions to prescribe long term hormone replacement therapy: results of a randomized controlled trial. *J. Gen Intern Med* 1999; **14**: 591–8.
51. Freidman PD, Brett AS, Mayo-Smith MF. Differences in generalists' and cardiologists' perceptions of cardiovascular risk and the outcomes of preventive therapy in cardiovascular disease. *Ann Intern Med* 1996; **124**: 414–21.
52. Forrow A, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med* 1992; **92**: 121–4.
53. Bobbio M, Brunella D, Guido G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994; **343**: 1209–11.
54. Bloor K, Freemantle N. Lessons from international experience in controlling pharmaceutical expenditure II: influencing doctors. *Br Med J* 1996; **312**: 1525–7.
55. Walley T, Wilson R, Bligh J. Current prescribing in primary care in the UK. Effects of the Indicative Prescribing Scheme and GP Fundholding. *PharmacoEconomics* 1995; **7**: 320–31.
56. Bateman DN, Campbell M, Donaldson LJ, Roberts SJ, Smith JM. A prescribing incentive scheme for non-fundholding general practices: an observational study. *Br Med J* 1996; **313**: 535–8.
57. Abel-Smith B, Grandjeat P. Pharmaceutical consumption. *Social Policy Series* 38. Brussels: Commission of the European Communities, 1978.
58. Feely J. The influence of pharmacoeconomic factors on prescribing patterns in Ireland. *PharmacoEconomics* 1992; **2**: 99–106.
59. Rittenhouse BE. Economic incentives and disincentives for efficient prescribing. *PharmacoEconomics* 1994; **6**: 222–32.
60. Seo T. Prescribing and dispensing of pharmaceuticals in Japan. *PharmacoEconomics* 1994; **6**: 95–102.
61. Barrett-Connor E. Post-menopausal estrogen and prevention bias. *Ann Intern Med* 1991; **115**: 455–6.
62. Lanza FL. Endoscopic studies of gastric and duodenal injury after the use of ibuprofen, aspirin, and other nonsteroidal anti-inflammatory agents. *Am J Med* 1984; **77** (1A): 19–24.
63. Garcia-Rodriguez LA, Walker LAM, Perez-Gutthann S. Nonsteroidal anti-inflammatory drugs and gastrointestinal hospitalisations in Saskatchewan: a cohort study. *Epidemiology* 1992; **3** (4): 337–42.
64. Langman MJS, Weil J, Wainright P. Risks of bleeding peptic ulcer associated with individual nonsteroidal anti-inflammatory drugs. *Lancet* 1994; **343**: 1075–8.
65. MacDonald TM, Morant SV, Robinson GC *et al.* Association of gastrointestinal toxicity of nonsteroidal anti-inflammatory drugs with continued exposure: a cohort study. *Br Med J* 1997; **315**: 1333–7.
66. Emery P, Zeidler H, Kvien TK *et al.* Celecoxib versus diclofenac in long-term management of rheumatoid arthritis: randomised double-blind comparison. *Lancet* 1999; **354**: 2106–11.
67. Simon LS, Weaver AL, Graham DY, *et al.* Anti-inflammatory and upper gastrointestinal effects of celecoxib in rheumatoid arthritis. *J Am Med Assoc* 1999; **282** (20): 1921–8.
68. Laine L, Harper S, Simon T, *et al.* For the Rofecoxib arthritis endoscopy study group. A randomised trial comparing the effect of rofecoxib, a cyclooxygenase 2-specific inhibitor, with that of ibuprofen on the gastroduodenal mucosa of patients with osteoarthritis. *Gastroenterology* 1999; **117**: 776–83.

5

When Should One Perform Pharmacoepidemiology Studies?

BRIAN L. STROM

Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine,
Philadelphia, PA, 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 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 of conducting the study itself. They 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 of these 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, clinical, and legal (see Table 5.1). Each will be of importance to different organizations and individuals involved in deciding whether to initiate a study. Any given study will usually be performed for several of these reasons. Each will be discussed individually.

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 re-positioning the drug
 - (a) Different outcomes, e.g., quality of life and economic
 - (b) Different types of patient, e.g., the elderly
 - (c) New indications
 - (d) Less restrictive labeling
 - (4) To protect the drug from accusations about adverse effects
- (C) Legal
- (1) In anticipation of future product liability litigation
- (D) Clinical
- (1) Hypothesis testing
 - (a) Problem hypothesized on the basis of drug structure
 - (b) Problem suspected on the basis of preclinical or premarketing human data
 - (c) Problem suspected on the basis of spontaneous reports
 - (d) Need to better quantitate the frequency of adverse reactions
 - (2) Hypothesis generating—need depends on whether
 - (a) it is a new chemical entity
 - (b) the safety profile of the class
 - (c) the relative safety of the drug within its class
 - (d) the formulation
 - (e) the disease to be treated, including
 - (i) its duration
 - (ii) its prevalence
 - (iii) its severity
 - (iv) whether alternative therapies are available
-

REGULATORY

Perhaps the clearest and most 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.¹ Many of these required studies have been randomized clinical trials, designed to clarify residual questions about a drug's efficacy. Others focus on questions of drug toxicity. Often it is unclear whether the pharmacoepidemiology study was a regulatory requirement or merely a "suggestion" by the regulator, but the effect is essentially the same. Examples of studies conducted to address regulatory questions include the "Phase IV" cohort studies performed of cimetidine² and prazosin.³ 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 a serious problem would be detected reliably and rapidly after marketing, it could feel more comfortable releasing the drug sooner. Although it is difficult to assess the impact of volunteered postmarketing studies on regulatory decisions, the very large economic impact of an earlier approval has motivated some manufacturers to initiate such studies. In addition, in recent years regulatory authorities have occasionally released a particularly important drug after essentially only Phase II testing, with the understanding that additional data would be gathered during postmarketing testing. For example, zidovudine was released for marketing after only limited testing, and only later was additional data gathered on both safety and efficacy, data which indicated, among other things, that the doses initially recommended were too large.⁴

Some postmarketing studies of drugs arise in response to case reports of adverse reactions reported to the regulatory agency. One response

to such a report might be to suggest a labeling change. Often a more appropriate response, clinically and commercially, would be to propose a 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 the increased risk of the disease is. As an example, a Medicaid database was used to study hypersensitivity reactions to tolmetin,⁵ following reports about this problem to the FDA's Spontaneous Reporting System.⁶

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

MARKETING

As will be discussed below, 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 clarify 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 under way 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 particularly likely to result from pharmacoepidemiology studies that recruit subjects for the study via prescribers. However, as discussed in Chapter 24, while this technique can be useful in selected situations, it is extremely expensive and less likely to be productive of scientifically useful information than most other alternatives available. In particular, the conduct of a purely marketing exercise under the guise of a postmarketing surveillance study, not designed to collect useful scientific information, is to be condemned.⁷ It is misleading and could endanger the performance of future scientifically useful studies, by resulting in prescribers who are disillusioned and, thereby, reluctant to participate in future studies.

Pharmacoepidemiology studies can also be useful to re-position a drug that is already on the market, i.e., to develop new markets for the drug. One could explore different types of outcome 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 35) and on patients' quality of life (see Chapter 36). 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 34), 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 a number of pharmacoepidemiology groups had data available on this question because of other studies which they had performed.⁸ McNeil was not as fortunate when questions were raised about anaphylactic reactions caused by zomepirac. If the data they eventually were able to have⁹ 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, and the risk, associated with the marketing of parenteral ketorolac, and chose to initiate a postmarketing surveillance cohort study at the time of the drug's launch.¹⁰⁻¹² Indeed, the drug was accused of multiple different adverse outcomes, and it was only the existence of this study, and later its 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 judgement 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 (see Chapter 9). A pharmaceutical manufacturer that is a defendant in such a suit cannot change whether the 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 above² was conducted because cimetidine was chemically related to metiamide, which had been removed from the market in Europe because it caused agranulocytosis. Alternatively, hypotheses can also be based on premarketing or postmarketing animal or clinical findings. For example, the hypotheses can come from spontaneous reports of adverse events experienced by patients taking the drug in question. The tolmetin,⁵ piroxicam,⁸ zomepirac,⁹ and ketorolac¹⁰⁻¹² questions mentioned above are all examples of this. Finally, an adverse effect may clearly be due to a drug, but a study may be needed to quantitate its frequency. An example would be the postmarketing surveillance study of prazosin, performed to quantitate the frequency of first dose syncope.³ Of course, the hypotheses to be tested can involve beneficial drug effects as well as

harmful drug effects, subject to some important methodologic limitations (see Chapter 34).

Hypothesis Generating

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

For example, it is generally agreed that new chemical entities are more in need of study than so-called “me too” drugs. This is because the lack of experience with related drugs makes it more likely that the 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. Previous experience with other drugs in the same class can be a useful predictor of what the experience with the 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 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 be used mainly in institutions, where there is close supervision, may be less likely to need such a study. 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 pre-marketing study. Drugs used to treat common diseases are important to study, as many patients are likely to be exposed to

them. Drugs used to treat mild or self-limited diseases 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 to differentiate safety, however, 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 a *degree of safety*. Specifically, a drug “is safe if its risks are judged to be acceptable.”¹⁴ Measuring risk is an objective but probabilistic pursuit. A judgement about safety is a personal and/or social value judgement about the acceptability of that risk. Thus, assessing safety requires two extremely different kinds of activity: measuring risk and judging the acceptability of those risks.¹⁴ 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 regulatory or legal problem. 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. There are a number of factors that can affect one's willingness to tolerate the risk of adverse effects from drugs (see Table 5.2). Some of these are related to the adverse outcome being studied. Others are related to the exposure and the setting in which it 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.

Critical 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 the excess risk. Tampons are

extraordinarily closely linked to toxic shock: the relative risk appears to be between 10 and 20. However, toxic shock is sufficiently uncommon that even a ten- to 20-fold increase in the risk of the disease still leads to an extraordinarily small risk of the toxic shock syndrome in those who use tampons.¹⁵

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 one of these diseases.

It is also material 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 that could make that delayed effect irrelevant or, at least, mitigate its impact. Thus, a delayed adverse event may be worth incurring if it can bring about beneficial effects today.

It is 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

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

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 of the absence of treatment. For example, the six year delay in introducing β -blockers into the US market has been blamed for resulting in more deaths than all recent adverse drug reactions combined.¹⁶ 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 that had been performed.⁴ Analogously, studies of toxic shock syndrome associated with the use of tampons were of public health importance, despite the infrequency of the disease, because consumers could then choose among tampons that were shown to carry different risks.¹⁵

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, nor 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.¹⁷

PERCEPTIONS OF THE EVALUATOR

Finally, much ultimately depends 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.¹⁸ 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 that 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. However, this subjective component is part of what makes informed consent so important. Most people feel that the final subjective judgement 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 judgement, it is useful to have some quantitative information about the risks inherent in some other activities. Some such information is presented in Table 5.3.

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

Hazard	Annual death rate (per 100 000 exposed individuals)
Heart disease (US, 1985)	261.4
Sport parachuting	190
Cancer (US, 1985)	170.5
Cigarette smoking (age 35)	167
Hang gliding (UK)	150
Motorcycling (US)	100
Power boat racing (US)	80
Cerebrovascular disease (US, 1985)	51.0
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 use (age 25–34)	4.3
Pedestrian (US)	3.8
Bicycling (US)	1.1
Tornados (US)	0.2
Lightning (US)	0.05

* Data derived from references 18–20.

CONCLUSIONS

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.

REFERENCES

1. 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.
2. Humphries TJ, Myerson RM, Gifford LM, *et al.* A unique postmarket outpatient surveillance program of cimetidine: report on phase II and final summary. *Am J Gastroenterol* 1984; **79**: 593–6.
3. Joint Commission on Prescription Drug Use. *Final Report*. Washington: Joint Commission on Prescription Drug Use, 1980.
4. Young FE. The role of the FDA in the effort against AIDS. *Public Health Rep* 1988; **103**: 242–5.
5. Strom BL, Carson JL, Schinnar R, *et al.* The effect of indication on the risk of hypersensitivity reactions associated with tolmetin sodium and other nonsteroidal antiinflammatory drugs. *J Rheumatol* 1988; **15**: 695–9.
6. Rossi AC, Knapp DE. Tolmetin-induced anaphylactoid reactions. *New Engl J Med* 1982; **307**: 499–500.

7. Strom BL, and members of the ASCPT Pharmacoepidemiology Section. Position paper on the use of purported postmarketing drug surveillance studies for promotional purposes. *Clin Pharmacol Ther* 1990; **48**: 598.
8. Bortnichak EA, Sachs RM. Piroxicam in recent epidemiologic studies. *Am J Med* 1986; **81**: 44–8.
9. Strom BL, Carson JL, Morse ML, *et al.* The effect of indication on hypersensitivity reactions associated with zomepirac sodium and other nonsteroidal antiinflammatory drugs. *Arthrit Rheum* 1987; **30**: 1142–8.
10. Strom BL, Berlin JA, Kinman JL, Spitz PW, Hennessy S, Feldman H, Kimmel S, Carson JL. Parenteral ketorolac and risk of gastrointestinal and operative site bleeding: a postmarketing surveillance study. *J Am Med Assoc* 1996; **275**: 376–82.
11. Feldman HI, Kinman JL, Berlin JA, Hennessy S, Kimmel SE, Farrar J, Carson JL, Strom BL. Parenteral ketorolac: the risk for acute renal failure. *Ann Intern Med* 1997; **126**: 193–9.
12. Hennessy S, Kinman JL, Berlin JA, Feldman HI, Carson JL, Kimmel SE, Farrar J, Harb G, Strom BL. Lack of hepatotoxic effects of parenteral ketorolac in the hospital setting. *Arch Intern Med* 1997; **157**: 2510–14.
13. Rogers AS, Porta M, Tilson HH. Guidelines for decision making in postmarketing surveillance of drugs. *J Clin Res Pharmacol* 1990; **4**: 241–51.
14. Lowrance WW. *Of Acceptable Risk*. Los Altos: William Kaufmann Inc., 1976.
15. Stallones RA. A review of the epidemiologic studies of toxic shock syndrome. *Ann Intern Med* 1982; **96**: 917–20.
16. Binns TB. Therapeutic risks in perspective. *Lancet* 1987; **2**: 208–9.
17. Marwick C. FDA ponders approaches to curbing adverse effects of drug used against cystic acne. *J Am Med Assoc* 1988; **259**: 3225.
18. Urquhart J, Heilmann K. *Risk Watch—the Odds of Life*. New York: Facts on File, 1984.
19. O'Brien B. “*What Are My Chances Doctor?*”—*a Review of Clinical Risks*. London: Office of Health Economics, 1986.
20. Silverberg E, Lubera JA. Cancer statistics. *CA Cancer J Clin* 1988; **38**: 5–22.

Part II

PERSPECTIVES ON PHARMACOEPIDEMOLOGY

6

The Public Health, the University, and Pharmacoepidemiology

PAUL D. STOLLEY

University of Maryland School of Medicine, Baltimore, MD, USA

JOAN-RAMON LAPORTE

Universitat Autònoma de Barcelona, Barcelona, Spain

INTRODUCTION

Pharmacoepidemiology is essentially a branch of public health, and, as such, it has as its prime goal the gathering of information leading to protecting the health of populations and improving the safety and efficacy of medications. Viewed from the perspective of academia, the field has as its goal the provision and evaluation of a public service. However, in a capitalist nation, the manufacture and sale of drugs are a branch of private sector economic activity and this fact can often lead to a conflict of goals.

The examples in Box 6.1 illustrate how far afield one can wander from the goals of pharmacoepidemiology as a public health activity, if one does not share these goals or simply views the field as an interesting quantitative exercise. Some practitioners are noted for ignoring the serious health consequences of adverse reactions or trivializing results and casting aspersions on findings as an exercise in debate (or for compensation as a well paid consultant). There is also the mistaken idea in

some quarters that pharmacoepidemiologists should act as plaintiff or defense lawyers and provide the client with the best defense, even if it means jeopardizing the public health.¹

It is our view that, to function as a responsible and ethical epidemiologist working in the field of adverse drug reactions, one must possess a knowledge of the history of drug development and regulation, be familiar with the epidemiology and the natural history of the disease under study, have a background in the economics of the drug industry, and understand the history of drug adulteration and of efforts to resist testing for safety and efficacy.

EFFICACY VERSUS EFFECTIVENESS: NEED AND AIMS OF PHARMACOEPIDEMIOLOGY

At the time a new drug is marketed, much information has been systematically obtained by means of preclinical and clinical studies aimed at defining its therapeutic role (see also Chapters 1

Aspirin and Reye's Syndrome

In the early 1980s, after several overwhelmingly convincing case-control studies indicted aspirin as a contributory cause of this disease, the aspirin manufacturers exerted political and economic pressure on the American Academy of Pediatrics to delay recommending that aspirin administration to children should be avoided during the winter months. At the same time there was political pressure on the US FDA to delay the implementation of warning labels on the aspirin packages. The over-the-counter analgesic market is large, lucrative, and competitive, and such warnings would interfere with sales and profits during the influenza season.

Benefit and risk of calcium channel blockers (CCBs) and therapeutic guidelines for the treatment of hypertension

Several clinical trials indicated that, compared to other antihypertensive treatments, CCBs were associated with an increase in the risk of death, ischemic heart disease, other cardiovascular events, and severe hemorrhage. These risks were also seen in several observational studies. Meta-analyses of clinical trials of CCBs in ischemic heart disease also suggested an increased risk of death.

Ironically, the risks seemed to be highest among hypertensive diabetic patients, one of the subgroups for which these drugs had been more heavily promoted without any proof of benefit and with the blessing of the US FDA and other national drug regulatory agencies. An investigation into possible conflicts of interest on this issue found a strong association between published positions on the safety of CCBs and the authors' financial relationships with pharmaceutical manufacturers. Following publication of these studies, guidelines for the treatment of hypertension were published which advocated the use of CCBs as first choice agents, almost interchangeable with low dose thiazide diuretics and β -adrenergic blocking agents. This recommendation was based on the results of the SYST-EUR trial, a placebo controlled study of uncertain quality using nitrendipine in the treatment of systolic hypertension in the elderly. At the end of the first year of this study, 44% of the patients initially allocated to nitrendipine were also on enalapril and/or hydrochlorothiazide, while patients initially allocated to placebo received more placebos in the case of uncontrolled hypertension. These guidelines have been judged as "consensus based," rather than "evidence based."

Box 6.1. Interests and conflicts of interest in pharmacoepidemiology.

and 4). After initial pharmacological screening using *in vitro* and *in vivo* models, pharmacodynamic, pharmacokinetic, and toxicological experiments are usually performed on animals. Pharmacodynamics is the study of the drug actions and effects on the body. Conversely, pharmacokinetics is the study of the actions of the body on the drug (drug disposition, including absorption, distribution, metabolism, and excretion). Toxicological studies include those after single and repeated doses, as well as special tests, including genotoxicity, mutagenesis, and carcinogenesis. If satisfactorily completed, this preclinical phase is followed by a clinical research and development program. This is arbitrarily divided into three premarketing phases.

Phase I is the first use of the new drug in human beings. Its main aim is the study of safety. It is generally undertaken in healthy volunteers, although certain categories of drugs (e.g., cancer

chemotherapeutic agents) are necessarily studied in real patients.

Phase II aims at defining the pharmacodynamics and the pharmacokinetics of the new drug in the human being. Its main objective is to define the range of doses that will be used in phase III studies and in usual clinical practice. It is performed in healthy volunteers or in patients, depending on the drug and its indication.

Phase III aims to establish the therapeutic value of the new drug, by comparing it with a placebo and/or with alternative drugs used for its potential indications. This is generally carried out by means of the randomized controlled clinical trial (RCT). Subjects participating in a clinical trial are randomly assigned to one of the treatments under comparison. The groups produced by randomization tend to have equivalent baseline known prognostic characteristics. It is assumed that unknown prognostic factors will also be equally

distributed across the different groups. Therefore, any difference between the groups at the end of the observation period can be attributed to the different treatments assigned to each group. Because randomization avoids various sources of potential confounding, the RCT is considered the “gold standard” for the evaluation of the efficacy of any therapeutic or diagnostic intervention, and it is a necessary and central step of the drug development process. Once RCTs with a new drug have been completed, application for drug registration is made by the manufacturer.

Despite the complexity of the steps of drug development, at its completion, knowledge of its future potential effects in practice is only partial. The information gathered by means of RCTs concern efficacy, i.e., the ability of the drug to bring about the intended effect under ideal conditions (e.g., in clinical trials). However, little is known about effectiveness (i.e., about the ability of the drug to bring about the intended effect in the usual clinical setting), because the circumstances of clinical use differ so markedly from those of clinical trials, in several respects.

Number of Patients

Clinical trials are performed in limited numbers of patients (see also Chapter 3). The number of individuals who have received a new drug at the time of its marketing is very limited, of the order of a few hundred to a few thousand. This precludes identifying adverse effects with an incidence of less than 1/100 or 1/1000. On the other hand, once the new drug is marketed, it may be taken by millions of people and rare adverse effects may appear that had not previously been identified. A serious adverse effect with an incidence of 1 in 10 000 or 1 in 100 000 patients may, for example, generate limited direct clinical notice: a single physician will rarely be able to prescribe the particular drug to a sufficient numbers of patients to ever see one case. Nevertheless, an adverse reaction even of this rarity may still have a substantial public health impact if the particular drug is taken every day by, say, 100 000 new patients and the reaction is life threatening.

Length of Exposure

Clinical trials are generally of relatively limited duration and, for many drugs, they are much shorter than the expected length of treatment in normal clinical practice. For example, a study of 80 clinical trials supporting the efficacy of five nonsteroidal anti-inflammatory drugs and two analgesic drugs introduced into the United Kingdom showed that in 25 studies treatment lasted for less than one day, in 15 for a minimum of 28 days, and only in four had the duration of treatment lasted for three months or longer. This implies that any beneficial or adverse effects that appear after a relatively long exposure to the drug are unlikely to be identified before marketing.

Representativeness of the Target Population

Participants in clinical trials are seldom representative of the general population. Phase I and phase II studies are usually performed in healthy volunteers and in patients who may not be representative of the future real users, with respect to susceptibility to the drug's effects and to drug disposition. Phase III clinical trials are generally performed in highly selected patients. The very young and the frail elderly are usually excluded. As a rule, selected participants have a single diagnosis; in contrast, the prevalence of one or more additional diseases among adults in primary health care is almost 40%. Patients with potential contraindications to the new drug and other high risk individuals are often excluded, thus tending to give a false impression of the benefit/risk ratio. They are usually healthier, younger, and of higher social status than the people who will ultimately receive the intervention. Patients in clinical trials have more accurate and clear cut diagnoses than those in routine practice (see Box 6.2). Thus, we believe that the deliberate and appropriate study of highly selected patients at the premarketing stage precludes the identification of those subgroups for whom the new drug may have an adverse risk-to-benefit ratio.

A meta-analysis of trials of antibiotic treatment in children with acute otitis media suggested that the treatment might not be effective. However, for ethical reasons, the sickest patients were excluded from many of the studies included in the meta-analysis. Therefore, it is difficult to generalize the results to all children with acute otitis media, including the sickest.

Several randomized clinical trials on thromboprophylaxis with oral anticoagulants in patients with atrial fibrillation have shown that these drugs significantly reduce the risk of stroke. However, fewer than 10% (3–40%, depending on the trial) of eligible patients with atrial fibrillation were finally included in these trials, thus hindering their interpretation and translation to therapeutic decisions.

In a recent meta-analysis of 25 trials of the treatment of congestive heart failure with β -adrenergic blocking agents, the mean age of patients in the trials was 60 years, 25% were female, 86% had already been treated with an ACE inhibitor, and left-ventricular dysfunction was a prerequisite for entry

into 24 of the 25 studies. In contrast, the average age of patients with heart failure in primary health care in the United Kingdom is 74 years, 54% are female, only 20% are treated with an ACE inhibitor, and yet only 30% have had ventricular imaging to measure ventricular dysfunction.

The summary estimates of discontinuation rates reported in clinical trials of antihyperlipidemic therapy were 31% for bile acid sequestrants, 4% for niacin, and 15% for gemfibrozil, while in cohorts of patients enrolled in two health maintenance organizations in the US, the one year probability of discontinuation was 41% for bile acid sequestrants, 46% for niacin, and 37% for gemfibrozil. Such discontinuation, of course, limits the effects of drug treatment.

While the discontinuation rates in long term clinical trials of antihypertensive drugs have been approximately 30%, dropout rates of approximately 50% have been reported in community based studies with one or more years of followup.

Box 6.2. Participants in clinical trials are seldom representative of the general population—some examples.

Disease Conditions Under Study

In developed countries, about 40% of all new disorders in primary health care do not evolve into conditions that meet accepted criteria for a diagnosis. The absence of clear definitions for undifferentiated disorders makes it extremely difficult methodologically to explore management using RCTs.

Drug Interactions

In phase III studies, the number of medicines that participants are allowed to take is generally limited to the drug under study or a small range of other drugs, thus precluding the identification of drug–drug interactions.

Dosage

Patients in clinical trials tend to follow precise protocol-driven dosage recommendations, while in actual community practice the doses of drugs tend to vary more widely, and thus drug inefficacy and

type A (dose dependent) adverse reactions are more likely.

Patient's Values Versus Clinical Research Endpoints

Drug trials tend to be driven by the mere existence of a new drug or a new potential indication, rather than by the need to answer clinically relevant questions that arise in practice. The majority of RCTs are promoted by drug manufacturers, as part of an investment that will yield a return once the new drug is marketed. Therefore, their main objective is to produce the evidence required by regulatory authorities, in order to apply for drug registration and licensing. If regulatory authorities do not require long term clinical outcome data in the broad population of patients in which the treatment is likely to be used, drug manufacturers are free to choose between performing definitive and expensive pragmatic clinical outcome studies mimicking routine clinical settings, or less expensive small studies using surrogate variables. According to

the FDA definition, a surrogate endpoint or marker is a laboratory measurement or physical sign that is expected to predict the effect of the therapy and is used in therapeutic trials as a substitute for a clinically meaningful endpoint that would be a direct measure of how a patient feels, functions, or survives. Long term outcome studies involve the risk of delayed marketing and a negative result, while a series of small studies aimed at pathophysiology, or on very selected patient populations, creates a less risky approach to product development.

Clinical Evaluation and Patient Followup

In a clinical trial, patient followup is likely to be more frequent and rigorous than in routine clinical practice, where clinical circumstances and judgment may dictate a rather different practice and patients may be less likely to follow the prescribed regimen or, for other reasons, less likely to receive the optimum therapy. It is then not surprising that, on average, patients in RCTs have better prospects of a favorable outcome than do patients treated in the community.

Evaluation of One Intervention Versus Various Simultaneous Interventions

The treatment of many chronic conditions may need more than one intervention. Unfortunately, however, clinical trials usually evaluate the effect of one intervention. Although specific designs exist to address this problem (e.g., factorial designs), clinical trials designed to evaluate two or more interventions simultaneously are scarce and difficult to conduct.

In summary, the generalization of results from trials is full of contradictions and complex conclusions. It seems clear that the observed differences should be regarded cautiously and should be carefully scrutinized in the real world of clinical practice. Generalization may be even more difficult for quality of life (see Chapter 36) and pharmaco-economic (see Chapter 35) assessments than for hard variables such as mortality.

DRUG SAFETY AND THE PUBLIC HEALTH

GLOBALIZATION, PHARMACOEPIDEMIOLOGY, AND DEVELOPING COUNTRIES—FROM WHO TO WTO

In each country, the final effectiveness of drugs depends on many factors: priorities in research and drug development in the pharmaceutical industry, registration and licensing policies and practices, local production, supply, quality control, distribution and management, priorities of the health care system, training of health care professionals, financial accessibility, dissemination of reliable information on drugs, general quality of medical care and of pharmaceutical dispensing, cultural perceptions and expectations, and the level of public education (see also Chapter 27).

The primary goal of health services regarding pharmaceuticals is to deliver efficacious drugs to patients in a timely, efficient, and economic manner. This is particularly difficult to accomplish in developing countries.² Although there is wide variability among them, in these countries the main limitations regarding health care are insufficient infrastructure, organization, equipment, and personnel. Resources are improperly utilized, a large proportion being allocated to the relatively small urban population and to hospitals. Drugs account for 30–40% of total health expenditures.² A substantial proportion of medicines are procured from abroad on an irregular basis, and drug availability is unreliable. Quality control is inefficient or nonexistent, and counterfeit, adulterated, or poor quality pharmaceuticals are not unusual, with tragic consequences, such as the epidemics of deaths from acute renal failure due to acetaminophen syrups containing diethylene glycol.^{3,4} There is a lack of necessary professional resources for the evaluation of new drugs submitted for registration. There is at least one sales representative for every three to ten physicians and generally there are no regulations controlling information on pharmaceuticals disseminated by the manufacturers. A substantial proportion of medicines have inadequate proof of efficacy, many are irrational fixed dose combinations of

several drugs, and some of them are outdated or withdrawn from their original manufacturer's countries. In 36 countries, less than 50% of the population has real access to health care and medicines.⁵ Prescription and dispensing practices often fall short of any acceptable standard, illiteracy rates are high, cultural perceptions of conventional medicine vary, and as a result, patient adherence to a therapeutic plan is unreliable. Self-medication with potentially dangerous over the counter and prescription drugs is common, partly due to what has been called the UTC ("under the counter") market, i.e., the sale of prescription only drugs without a prescription.^{6,7} Clinical follow up is rare, and adverse effect management and reporting are generally nonexistent or unknown to health professionals. All these circumstances add to the gap between clinical trials and usual practice described in the previous section. Consumption and drug utilization statistics are generally unavailable or kept secret.

In the 1970s, the World Health Organization's Action Programme on Essential Drugs (ED) signaled a global recognition that a limited list of drugs could meet the health needs of most of the population in developing countries.⁸ The objective of the WHO's Action Programme on ED was to ensure the availability of EDs to all members of the community at the lowest possible cost, by encouraging countries to set priorities for drug use and purchasing, dissemination of reliable drug information for health professionals and the public, and regional collaboration in bargaining with the multinational firms. Technical cooperation started in a number of countries to develop lists for their use at the different levels of the health care system. The program was set in the context of the overall objective of Health For All (HFA) By The Year 2000. Assistance was given to countries in the formulation of ED policies and strategies. Cooperation was established with other UN agencies. Since then, and despite initial resistance,⁹ the ED concept has been successfully adopted in a number of countries. Its implementation, nevertheless, has been generally limited to the public sector.⁵ The ED concept is evidence based, is simple, promotes equity, and is rooted in firm public health principles.¹⁰

At the beginning of the new millennium the situation is changing. Today, new players, such as

the World Bank and, increasingly, the World Trade Organization (WTO), have an influence on international health.^{11,12} Current world economic trends favor a strategy of development based on the liberalization of markets and on the assumption that free flow of trade, finance, and information will produce the best possible outcome for economic development. Globalization of the world economy is not only integrating trade, finance, and investment, but also consumer markets, and in particular the pharmaceutical market, with dramatic consequences on prices (unique worldwide prices may be set), therefore jeopardizing access to drugs.

Globalization has produced a cultural shift from the equitable Health For All principle, to the more pragmatic and market oriented "global burden of disease" (GBD) concept, by which the world is divided into market areas, where the basic assumption is that health indicators, such as life expectancy and disability from communicable and noncommunicable diseases, as well as their projections for the year 2020, are expressed in terms of costs.¹³⁻¹⁶ These costs are affordable only for certain populations living in wealthier countries and a minority living in less developed countries. The global burden of disease data reflect profound global inequalities (see Box 6.3).^{17,18}

On the other hand, the "demographic transition," i.e., rapid growth of population and decrease in birth rates due to improved living and health conditions, leading to aging of the population, produces a shift from acute to chronic conditions, with the resulting increased health needs and health care costs.⁸

Four processes have contributed to specifically shape this new situation created by globalization in the field of pharmaceuticals. These are the TRIPs Agreement, health sector reform and liberalization, moves to closer cooperation and harmonization of procedures among drug regulatory authorities, and mergers of pharmaceutical firms.

Patents: The WTO TRIPs Agreement

The Uruguay Round of negotiations on multi-lateral trade led to the creation of the World Trade Organization (WTO) in 1995. Its purposes are to help in the development of trade in an open

98% of all deaths in children younger than 15 years occur in the developing world. 83% of deaths at 15–59 occur in the developing world. The probability of death between birth and 15 years ranges from 22% in sub-Saharan Africa to 1.1% in the established market economies.¹³ Developed regions account for 11.6% of the worldwide burden from all causes of death and disability, and for 90.2% of health expenditure worldwide.¹⁵

The following are per capita health and drug expenditures by global region in 1990. (Bennett *et al.*, quoted)⁵

Region Public	Number of countries (%)	Health expenditures		Pharmaceutical expenditures		
		Total <i>per capita</i> (\$)	As % of GDP	Total <i>per capita</i> (\$)	As % of GDP	% in public sector
Established market economies	25	1675	7.73	137.5	0.6	59.8
Middle Eastern Crescent	32	189	4.27	26.8	0.7	26.0
Former Socialist countries	19	150	4.27	19.5	–	–
Latin America and Caribbean	33	118	5.30	26.4	0.9	28.5
Asia and Pacific islands	33	60	4.01	11.8	0.6	18.6
Sub-Saharan Africa	47	36	4.86	7.8	0.9	33.2

Recent figures from WHO indicate that one-third of the world's population does not have regular access to essential drugs.¹⁷ Pharmaceutical consumption in the top ten countries (including 17.5% of total world population) in the year ending 30 June, 1999 (amounting to 190 billion dollars) was around 60% of total world consumption in that year (amounting to 315 billion dollars). Developing countries, where three-quarters of the world's population live, produce less than 10% of total global pharmaceutical production and account for less than 25% of annual global expenditure on drugs.

Although in most developing countries expenditure on drugs doubles every four years, the gross national product doubles only every 16 years.² The WHO estimates that of the 56 billion dollars spent each year on health research, less than 10% is spent on diseases known to affect 90% of the global population.

The following are examples of annual *per capita* incomes of less developed countries and costs of health technologies:¹⁸

Country	Annual <i>per capita</i> Income (\$)	Health technology	Cost (\$)
Tanzania	120	Ceftriaxone, 7-day course	130
Haiti	250	Streptokinase, one course	400
Egypt	790	One course of treatment for multidrug resistant tuberculosis	5 400
Barbados	6560	HIV triple therapy (one year course)	16 000

Box 6.3. Global inequalities in health.

system, to settle trade disputes between governments, to organize trade negotiations, and to supervise global trade agreements approved during the “Uruguay Round”.¹⁹ One of these is the Trade-Related Intellectual Property Rights (TRIPs) Agreement, which links intellectual property and trade issues for the first time and establishes minimum universal standards for trade. Pharmaceuticals are treated like any other technological product, insofar as the granting of patent

protection is concerned. Under the TRIPs Agreement, all Member States have to make patent protection available for at least 20 years to any invention of a pharmaceutical product for which a patent application had been filed after 1 January 1995.^{20,21} This is likely to have an important effect on the equitable access of populations to health and drugs, especially in developing countries (see Box 6.4).^{9,10,22–25} Case studies in Asian countries and in Argentina suggested that prices will increase,

A shift has occurred from the human rights oriented WHO's "Health for All" principle to the economic and market oriented World Bank's health sector reform and WTO's patent protection policy.

The future availability and accessibility of pharmaceuticals will be significantly affected by patent protection (as a result of the WTO Agreement on TRIPs), economic integration (with the establishment of free-trade zones which require harmonization to facilitate the free flow of goods between countries),²³ health sector reform (with decentralization and privatization), deregulation of pharmaceuticals (expedited approval and lifting of price controls), and mergers of pharmaceutical companies.

Far from the promised transfer of technology, since the introduction of protection for pharmaceuticals, 11 Argentinean companies have been acquired by foreign multinationals, and a dozen formulation plants have been closed down in the Andean countries. The prices of patented and nonpatented drugs and drug bills are expected to rise because of the lack of competition resulting from the disappearance of local laboratories producing cheap, nonpatented drugs. In Argentina, this increase may be more than \$1 billion per year.²⁴

Mergers of pharmaceutical companies give rise to huge transnational corporations with increasing

influence. In 1999, the top 10 pharmaceutical companies had a business volume of 113 billion dollars, equivalent to 36% of the global market, and the top 20 had a business volume of 168 billion dollars (53.5% of the total world market).

Liberalization of pharmaceutical trade can result in real concentration of production in certain countries, so that multinational firms will be free to export finished or semifinished products rather than transferring technology or foreign investment directly to developing countries.

Of 1223 new chemical entities launched between 1975 and 1997, 379 (30.9%) were therapeutic innovations, but only 11 (1%) were specifically designed for tropical diseases. Most such products were either incidental discoveries recovered from veterinary medicine or molecules discovered by governmental or academic institutions and only later acquired and sold by the pharmaceutical industry. The development of eflornithine, a new drug for "sleeping sickness", was discontinued because it was unprofitable. Two drugs for use in "tropical diseases" have been produced since the TRIPs agreement in 1994, and they are updated versions of existing products (new formulations of pentamidine and amphotericin B).^{2,9,10}

Box 6.4. Cultural, social, and economic effects of globalization in less developed countries.^{2,9,10}

the local rate of pharmaceutical research and development and the flow of technology and investments to developing countries is not likely to increase, research on diseases prevalent in developing countries will not be promoted, and local pharmaceutical industry will be dismantled.^{26,27} Developing and least developed countries have been granted a period of grace of five to 10, or 11 years, depending on their level of development, in which to amend their national legislation in accordance with the standards of the TRIPs Agreement.²⁷ The recent use of TRIPs to argue that access to the full clinical trial data would breach the intellectual property rights of an applicant for a marketing license in Europe, suggests that trade agreements reinforce justifications for commercial secrecy to the detriment of scientific openness and drug regulation.²⁷

Prior to the TRIPs Agreement, many developing countries did not make patent protection available

for pharmaceuticals, in order to allow the manufacture of copies and generic equivalents of drugs by local manufacturers at reduced prices. Such nonpatent regulation helped some countries (namely the Republic of Korea, Egypt, Argentina, Brazil, and Mexico) to build an indigenous pharmaceutical industry based on imitative cheaper drugs, in the same way as, during the 1970s and the 1980s, other countries in Western Europe and Japan persistently resisted providing pharmaceutical product patents for the same reasons.²²

On the other hand, implementation of the WTO's principles of free circulation of goods should imply that the procedures, criteria, and drug regulations of more developed countries are implemented just as rigorously in developing countries. However, poor social development prevents such implementation.²⁸

The majority of drugs included in limited lists for the public sector in developing countries have

expired patents. However, new essential drugs are needed to meet pressing public health needs, and not all are affordable, or available due to the prohibitive prices that result from patent protection (see below), fluctuating international production, waste of resources due to counterfeiting or substandard production, or unsuitability for use in field conditions. An international coalition of health and legal professionals, including *Médecins Sans Frontières*, has recently warned that patent protection set up through the TRIPs Agreement is preventing poor countries from producing cheaper local versions of essential drugs.^{2,8,10} Examples are azithromycin (trachoma), ceftriaxone (resistant bacterial meningitis), ciprofloxacin (resistant *Shigella* dysentery), didanosine, indinavir, lamivudine, nevirapine, and zidovudine (HIV infection), fluconazole (fungal infections), and ofloxacin (multidrug resistant tuberculosis).²⁹ Lack of access to essential drugs and vaccines because of economic factors raises human rights issues. However, it is evident that financial access to pharmaceuticals does not necessarily imply effective use if it is not linked to the correction of a wide range of other deficiencies. Development of an effective national drug policy, adoption of the ED concept, use of compulsory licensing, parallel imports, promotion of the production and use of generic drugs, improvement of pharmaceutical management, dissemination of reliable drug information, and continuous training for health professionals are all measures that have been proposed to mitigate the effects of TRIPs on public health.^{2,10,27}

STRUCTURAL ADJUSTMENT, HEALTH SECTOR REFORM AND DEREGULATION

Economic adjustment introduced by the International Monetary Fund and the World Bank includes currency devaluation and liberalizing trade by increasing exports, decreasing imports, and increasing local access for foreign and transnational corporations. Government spending is cut by introducing user charges, withdrawing subsidies, privatization, retrenching civil servants, and introducing wage restraints.³⁰ According to Unicef, during the 1980s structural adjustment has produced a drop of 10–25% in average incomes, a

25% reduction in *per capita* spending on health, and a 50% reduction in *per capita* spending on education in the poorest countries of the world.³¹ A major change has occurred in the roles of WHO and the World Bank in the health sector. By applying a strategy of loan funding, the World Bank advocates health sector reform, including decentralization, privatization, and cost recovery.^{32–34} In some countries, the World Bank has become active in funding the purchase of pharmaceuticals. The probity of financing a recurrent expenditure with loan funds has been challenged.⁸

The imperative to liberalize trade is reducing state involvement in the social sectors. Health sector reform involves not only privatization processes,³⁵ but also deregulation of drugs,²³ with potentially dramatic consequences on populations with substantial illiteracy rates and low educational levels, living in countries where independent information is scarce or non-existent.³⁵

DRUG REGULATION AND HARMONIZATION

The aim of drug regulation is protection of the public (see also Chapter 8). The choices involved in deciding which medicines are useful and safe are complex and need knowledge, skill, and experience. Clinical and epidemiological drug safety experience show that the consequences of licensing ineffective or unsafe drugs can be very serious. Thus, drugs treated as mere consumer goods become a danger to public health. The broad responsibilities of the state with regard to pharmaceuticals should at least include registration, quality control, monitoring of drug use (including supervision of production, imports, and distribution as well as sale at pharmacies), monitoring of adverse drug reactions and of drug prescription and use, and informing and educating health professionals and the public.

Registration is no more than a step in the regulatory process, although it is one of the most important. In the United States, the FDA has the legal mandate to review all new drugs for licensing for sale. The manufacturer must prove the drug is “safe and effective” for the medical indications claimed. The legal requirement for efficacy is

relatively recent. The amendment to the law establishing this requirement was enacted in 1964, but the legal definition of “efficacy” was settled only after litigation and a 1971 Federal Court ruling. In most instances, the generally accepted legal requirement for proof of “efficacy” is a well designed, carefully conducted, and fully reported randomized controlled trial. Once a drug is licensed, there is no legal requirement to monitor it or keep it under formal surveillance for unexpected or rare side-effects, other than the spontaneous reporting system (see Chapters 8, 10, and 11), but the FDA has managed to persuade many manufacturers to establish such a postmarketing surveillance system “voluntarily”, using persuasion, or the tacit threat of slower or nonapproval. Often manufacturers understand such surveillance is ultimately in their own best interests (see Chapter 7).

Drug regulatory authorities have certain manifest obligations towards international collaboration, such as the need to eliminate counterfeiting and substandard products, to assure good manufacturing practices, and other aspects of control. Earlier international cooperation initiatives, such as the development of International Non-proprietary Names (INN) and collaboration in adverse drug reaction monitoring led by WHO, have been significant improvements in drug management and exchange of information.^{36,37}

During the 1990s, the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) has had a central role in globalizing the pharmaceutical market. The ICH was established to function as a forum to harmonize technical requirements for registration of new innovative pharmaceuticals. Its goals were to reduce redundancy and costs of drug development while hastening the delivery of proven new therapeutic agents to patients.^{37,38} It has set defined technical specifications regarding the testing of drug quality, safety, efficacy, and communication among authorities.³⁹ The ICH will certainly have much wider reaching implications on the marketing, regulation, and utilization of drugs than has ever been explicitly stated.³⁹ It was convened by the regulatory authorities of the US, Japan, and the European Union (EU), the secretariat is provided

by the International Federation of Pharmaceutical Manufacturers Associations (IFPMA), WHO is an observer, and developing country regulators and generic manufacturers have been excluded. The composition and membership has created concern.^{8,37} It has been criticized for being too much concerned with getting new drugs onto the market as quickly as possible, while paying insufficient attention to monitoring patterns of use, drug safety, and drug information, and for lacking accountability and openness.³⁷ Its guidelines will benefit larger companies, which will now be able to submit the same set of data to regulatory authorities worldwide, but they will probably marginalize medium- and small-sized manufacturers in developing countries, which will not be able to meet the technical requirements outlined by ICH. The US Generics Pharmaceutical Industry Association (GPIA) noted that the requirements for generics were being increased without a scientific need to do so.³⁷

In the EU, harmonization has mainly focused on the standardization of the criteria for new drug approval in member states and in developing common procedures for drug licensing and for safety review. Other aspects, such as a review of existing marketed products and the adoption of common standards governing information, advertising, and marketing of drugs have been omitted, sometimes because of strong opposition from the drug industry.⁴⁰ Any inclusion of a “need clause” in the criteria for drug registration (i.e., that the manufacturer not only has to establish that a drug is safe and effective but also that there is a need for that drug—in other words that it has advantages over other already existing drugs) would constitute a trade barrier, in contravention of the Treaty of Rome, and is expressly excluded by the first pharmaceutical Directive (65/65/EEC).⁴¹ These steps are basically coherent with the fact that the European Medicines Agency functionally depends on the Industry Directorate of the European Union Commission, rather than the Consumers Protection Directorate (the EU has no Directorate on Health).

In developing countries, poor funding of drug regulatory agencies limits their activities mainly to registration and to routine controls (at best), and

generally monitoring and information activities are not undertaken. For this and other reasons, regulatory agencies in developing countries depend heavily upon the experience gained and the assessments carried out by larger national agencies elsewhere.⁷ The registration process is often feeble or simply nonexistent. In certain countries, “automatic registration” of drugs has been cynically advocated as a means to prevent corruption among drug regulators. In other countries the term “homologation” means that drugs are automatically licensed if they have been previously approved in a major country, with little or no local regard to manufacturing practices, quality control, need, approved indications, or information which will be disseminated by the manufacturer. These trends are the result of the lack of attention paid by governments to the role of regulation in

protecting the public health. In some countries, the legal functions of drug regulatory bodies have been limited by new legislation. Some worrying effects of these policies have been identified, such as registration of many ineffective or unnecessary drugs,⁴² and a decrease in consumption volume, despite an increase in market value, due to price increase.^{42,43}

MERGERS OF PHARMACEUTICAL FIRMS

There are certain features of the modern drug industry that affect the safety and efficacy of the drug supply. Like so many aspects of the economy, the drug industry is global in nature and is characterized by economists as an oligopoly. An oligopoly is a market dominated by a relatively small number of large firms, and with the mergers

Market success of pharmaceutical companies depends on the continuous launching of new products. Drug development costs are high and increasing, and only large companies can afford the investments necessary.

The proportion of revenue allocated to research and development in the larger pharmaceutical companies is around 15%. At Glaxo Wellcome, for example, marketing and general administrative expenditure accounts for more than twice this amount. With mergers, companies’ activities can be streamlined and duplication reduced.

Manufacturing plants can be closed down and production concentrated into fewer centers, in countries with cheaper labor. Locating production plants in particular countries may facilitate the introduction of the companies’ products in those countries. Similarly, the threat of closing the manufacturing plant may be used to obtain product registration and high prices. The power and influence of pharmaceutical companies thus increases. As the objective is maximizing profits, and it is unlikely that profit oriented organizations will devote much of their effort to poorer populations, large innovative firms focus on the most profitable segments of the market.

The losers are people dismissed from their jobs. It is estimated that the Glaxo Wellcome merger led to 8000 lost jobs. Stockbrokers and stockholders are the obvious winners. The following are changes in

market capitalization following announcement of the proposed merger of Glaxo Wellcome and SmithKline Beecham on 31 January 1998.

Company	Capitalization (\$billion)	
	31 Dec 1997	3 Feb 1998
Merck	121	139
Novartis	104	110
Roche Holdings	86	NA
Bristol-Myers Squibb	82	99
Glaxo Wellcome	80	113
Pfizer	75	107
SmithKline Beecham	53	76

Overconcentration may interfere with innovative activity and lead to monopolistic power. A monopoly of industries from the same (industrialized) country may not be so at the international level. As it is likely to benefit the national economy, and although close scrutiny of merger activity is important, there is little incentive for the governments of the country of origin to oppose any merger of two companies from that country, in the “national interest”. Under such circumstances, patients and governments of less fortunate countries have little to say.

of the past few decades, manufacturing capability is increasingly in the hands of fewer and larger firms with international operations (see Boxes 6.4 and 6.5).⁴⁴ An oligopoly confers distinct advantages upon the participating firms: economies of scale, less competition, opportunity for collusion with competitors on price and market share division, and ability to carry out expensive research and development. The sheer size of the firm along with monopolies due to patent protection permit price setting with little possibility of negotiation by the purchaser, be it patient, hospital, or national agency. The admission in 1999 of price fixing by seven of the largest firms manufacturing vitamins is an example of the societal costs of oligopolistic arrangements; 80% of the vitamin market was controlled by seven firms acting in illegal collusion.

The large multinational drug manufacturing firms have a capability to invest in research and development impossible for smaller ones. Utilizing their own laboratory facilities or purchasing the development "rights" to the research of university laboratories, drugs can move relatively quickly from discovery to development to marketing. The "downside" of these arrangements is the pressure for secrecy, which is in opposition to the usual ethos of scientific investigation and cooperation.

The multinational firms, acting in unison, can even "blackmail" an entire government, or at least attempt to do so. This is illustrated by the recent threat of Glaxo to withdraw their manufacturing plants from the United Kingdom, if the National Health Service refused to purchase their anti-influenza drug called "Relenza".

The potential ethical problems that can occur when academics accept industry funding are epitomized in the search for specific disease causing genes. On 18 October, 1999, the *Wall Street Journal* reported that the Glaxo Corporation had succeeded in identifying genes that are responsible for familial migraine, psoriasis, and type 2 diabetes mellitus. However, in order to protect Glaxo's rights to "patent" these genes, this information will be kept secret and not published for at least one year. Part of the work related to the gene search was performed at Duke University, which presumably agreed to the secrecy/nonpublication policy.

This nonsharing and secrecy surrounding research results is in direct opposition to traditional practice in academia and has profound practical and ethical implications. We believe that secret medical research contravenes conventional ethical codes and is contrary to the entire purpose of academia. Further, the policy of placing future profits ahead of collegial sharing of data is a dangerous precedent for public health.

SURVEILLANCE OF PRESCRIBING PRACTICES

Many chapters of this book deal with postmarketing surveillance of drugs for adverse effects, and this is indeed a most important function of pharmacoepidemiology. A less recognized, but increasingly relevant activity of pharmacoepidemiology is monitoring the prescribing patterns of physicians and national drug utilization (see also Chapters 30 and 31). Numerous studies show that even in the developed world, where reliable prescribing information is usually available (but not always consulted), prescribing patterns leave room for much improvement. Common suboptimal practices include prescribing antibiotics for self-limited viral respiratory syndromes, overprescribing of minor and major tranquilizers, prescribing of expensive and unnecessary combination products for the common cold or headache, and, paradoxically, underprescribing of certain effective agents, such as β -blockers and antiplatelet drugs after myocardial infarction, antihypertensive agents, antithrombotics for atrial fibrillation and analgesics for postoperative pain.⁴⁵⁻⁴⁸

National drug utilization studies consistently reveal extraordinarily high rates of use of tranquilizers, antibiotics, vitamins, and hypnotics. An older study of vitamin B₁₂ injection in England and Wales suggested a greater than 20-fold difference between the amount injected and the highest possible estimate of pernicious anemia prevalence (the only justified use, other than as a placebo).⁴⁹ This suggests either a misunderstanding of the vitamin's true indications or high use of it as a "placebo". Patient demand may also play a role here. Indiscriminant prescribing of antibiotics has

become an important problem all over the world. In the USA, the Centers for Disease Control have initiated a special program to educate both physicians and patients about the prudent use of these agents. Resistant strains of many bacteria have developed, in part due to overuse and inappropriate use of antibiotics.

Other problem areas in possible overutilization include the use of tranquilizers and the stimulant methylphenidate (Ritalin) indicated for “attention deficit disorder” (ADD), a somewhat vaguely and casually defined behavioral syndrome of school-age children. Estimates of the prevalence of Ritalin use in US children vary, but probably no less than one million children in the USA are daily users of this controversial but generally accepted pharmacotherapy. Long-term consequences of such use are under active investigation. While the drug is prescribed in other developed nations, the popularity of Ritalin treatment elsewhere does not come even close to rivaling the high utilization in the United States, where the ADD diagnosis is currently in vogue. No standardized surveillance system has been established, in spite of the high use rates and long term pattern of use of this amphetamine-like drug.⁵⁰

In summary, pharmacoepidemiologists have developed methods for tracking prescribing practices and national or regional drug utilization patterns but it is the rare drug regulatory agency that monitors and analyzes these patterns and then follows with regulatory or remedial action.

THE ROLE OF THE UNIVERSITY SCIENTIST IN THE FIELD OF PHARMACOEPIDEMIOLOGY

The role of the university scientist should reflect the unique role and function of the university: the provision of disinterested inquiry and critique. This role becomes a more difficult one to fulfill as university scientists become more and more dependent on funding from sources that have a vested interest in obtaining particular outcomes from the studies they fund. This problem confronts the scientist who receives funding from public as well as private sources.

Pharmacoepidemiologists are particularly at risk of becoming the targets of attempts to influence their research results, because the outcome of their studies can often affect the sales or even the licensing of drugs and medical devices. There are lamentable instances where epidemiologic studies have been conceived, designed, and conducted to obtain a predetermined result in order to obscure a drug/disease association, or, more rarely, to produce such an association.

It is still possible, but increasingly difficult, to defend the core values of academia and carry out contractual work with the private sector or certain government agencies subject to political pressures.

Yet another challenge to scientific integrity is the size in interest in homeopathic remedies. In Europe, homeopathic remedies have a large following and their popularity is increasing in the USA. Since the scientific rationale for the use of such “drugs” is completely absent, there is hardly any argument for even testing such concoctions, and yet the United Kingdom’s National Health Service has established a Royal Hospital of Homeopathic Medicine and in the USA academicians accept grants from government and the private sector to study those “medicines.” Academic medical centers, in an attempt to capture the market, to appear responsive to public demand, and to receive research grants, wind up pandering to the latest vogue in unscientific therapeutics under the pretense of doing a public service. In the process, they divert precious resources to predictably worthless activities.

CONCLUSIONS

We have identified major threats to the integrity and mission of pharmacoepidemiology as conducted in academic settings:

- increasing secret work and non-sharing of data due to contractual obligations with the private sector;
- diversion of resources to “fringe” medicine where no basis for rational study investigation has been presented;

- in the desire to obtain funding, conducting studies where a predetermined outcome is desired and engineered as a service to the funding agency;
- neglect by pharmacoepidemiologists of the plight of Third World consumers, where drug markets are often unregulated and drug advertising grossly and dangerously misleading; and
- increasing control and setting of research priorities in pharmacoepidemiology by the private sector to further its needs, which are not always congruent with public health goals.

We hope that by highlighting these threats, we can help protect against them.

REFERENCES

1. Stolley PD. Faith, evidence, and the epidemiologist. *J Public Health Policy* 1985; **6**: 37–43.
2. Bapna JS, Tripathi CD, Tekur U. Drug utilisation patterns in the Third World. *Pharmacoeconomics* 1996; **9**: 286–294.
3. Woolf AD. The Haitian diethylene glycol poisoning tragedy. A dark wood revisited. *J Am Med Assoc* 1998; **279**:1215–6.
4. O'Brien KL, Selanikio JD, Hecdivert C, Placide M-F, Louis M, Barr DB, Hospedales J, Lewis MJ, Schwartz B, Philen RM, St Victor S, Espindola J, Needham LL, Denerville K, for the Acute Renal Failure Investigation Team. Epidemic of pediatric deaths from acute renal failure caused by diethylene glycol poisoning. *J Am Med Assoc* 1998; **279**: 1175–80.
5. Bennett S, Quick JD, Velásquez G. Public-private roles in the pharmaceutical sector. Implications for equitable access and rational drug use. *Health Economics and Drugs*, DAP Series No. 5. World Health Organization Geneva: Action Programme on Essential Drugs, 1997.
6. Drug Utilization Research Group, Latin America. Multicenter study on self-medication and self-prescription in six Latin American countries. *Clin Pharmacol Ther* 1997; **61**:488–93.
7. Weerasuriya K. Globalisation of drug regulation and drug regulators in developing countries. *Int J Risk Safe Med* 1996; **9**: 187–193.
8. Laing R. The world health and drug situation. *Int J Risk Safe Med* 1999; **12**: 51–7.
9. Unwin N, Alberti G, Aspray T, Edwards R, Mbanya J-C, Sobngwi E, Mugusi F, Rashid S, Setel P, Whiting D. Economic globalisation and its effect on health. *Br Med J* 1998; **316**: 1401–2.
10. Pécoul B, Chirac P, Trouiller P, Pinel J. Access to essential drugs in poor countries. A lost battle? *J Am Med Assoc* 1999; **281**: 361–7.
11. Anonymous. The World Bank, listening and learning. *Lancet* 1996; **347**: 411.
12. Walt G. Globalisation of international health. *Lancet* 1998; **351**: 434–7.
13. Murray CJL, Lopez AD. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* 1997; **349**: 1269–76.
14. Murray CJL, Lopez AD. Regional patterns of disability—free life expectancy, and disability-adjusted life-expectancy: Global Burden of Disease Study. *Lancet* 1997; **349**: 1347–52.
15. Murray CJL, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet* 1997; **349**: 1436–42.
16. Murray CJL, Lopez AD. Alternative projections of mortality and disability by cause: 1990–2020: Global Burden of Disease Study. *Lancet* 1997; **349**: 1498–1504.
17. Anonymous. Progress of WHO Member States in developing national drug policies and in revising essential drugs lists. WHO/DAP/98.7. Action Programme on Essential Drugs. World Health Organization. Geneva, 1998.
18. Donald A. Technology transfer. The problem with “trickle down” theory. *Br Med J* 1999; **319**: 1298–9.
19. Velásquez G, Boulet P. Essential drugs in the new international economic environment. *Bull WHO* 1999; **77**: 288–92.
20. Otten A. Pharmaceutical patents and the TRIPS Agreement. *WHO Drug Information* 1998; **12**: 211–3.
21. Velásquez G, Boulet P. Globalization and access to drugs. Implications of the WTO/TRIPS Agreement. *Health Economics and Drugs*, DAP Series No. 7. Action Programme on Essential Drugs, Geneva: World Health Organization, 1997.
22. Bala K. Impact of WTO on national drug policies. *HAINews* 1998; **103**:1–2, 11–12.
23. Anonymous. Action Programme on Essential Drugs. Report of the biennium 96–97. WHO/DAP/MAC (10) 98.4. World Health Organization. Geneva, 1998.
24. Iglesias Rogers G. Reform of health care moves slowly in Latin America. *Lancet* 1999; **354**: 1104.
25. Trouiller P, Olliaro PL. Drug development output: What proportion for tropical diseases? *Lancet* 1999; **354**: 164.
26. Correa CM. South Centre. *WHO Drug Information* 1998; **12**: 215–6.
27. Mirza Z. Health Action International. *WHO Drug Information* 1998; **12**: 216–8.
28. Figueras A, Estévez F, Laporte JR. New drugs, new adverse reactions, and bibliographic databases. *Lancet* 1999; **353**:1447–8.

29. Yamey G. Agencies urge end to global trade restrictions on essential medicines. *Br Med J* 1999; **319**: 1455.
30. Wakhweya AM. Structural adjustment and health. *Br Med J* 1995; **311**: 71–2.
31. Abbasi K. The World Bank and world health. Under fire. *Br Med J* 1999; **318**: 1003–6.
32. Abbasi K. The World Bank and world health. Interview with Richard Feachem. *Br Med J* 1999; **318**: 1206–8.
33. Abbasi K. The World Bank and world health. Changing sides. *Br Med J* 1999; **318**: 865–9.
34. Abbasi K. The World Bank and world health. Healthcare strategy. *Br Med J* 1999; **318**: 933–6.
35. Stocker K, Waitzkin H, Iriart C. The export of managed care to Latin America. *New Engl J Med* 1999; **340**: 1131–6.
36. Edwards IR, Biriell C. Harmonisation in pharmacovigilance. *Drug Safety* 1994; **10**: 93–102.
37. Hodgkin C. International harmonisation—The need for transparency. *Int J Risk Safe Med* 1996; **9**: 195–199.
38. Goldsmith DI. The impact of International Conference on Harmonization (ICH) on pharmacoepidemiology and pharmacovigilance. *Pharmacoepidemiol Drug Safety* 1998; **7**: 289–291.
39. Baber N. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use (ICH). *Br J Clin Pharmacol* 1994; **37**: 401–4.
40. Anonymous. MINE for information. *Scrip* 1998; **235**: 4.
41. Hodgkin C. 1992—is harmonisation healthy? *ISDB Rev* 1992; **1**: 29–32.
42. Fefer E. Drug regulatory trends in Latin America. Presentation made at Latin America Health Care Congress, Nassau, 1998.
43. MacInnes A. Patents force issues in South America. *Scrip Mag* 1998; **72**: 48–52.
44. Wan Po AL. Mega-mergers in the pharmaceutical industry. In whose interests? *Pharmacoeconomics* 1998; **14**: 349–55.
45. Woods KL, Ketley D, Lowy A, Agustí A, Hagn C, Kala R, Karatzas NB, Leizorowicz A, Reikvam A, Schilling J, Seabra-Gomes R, Vasilauskas D, Wilhelmssen L. Beta-blockers and antithrombotic treatment for secondary prevention after acute myocardial infarction. Towards an understanding of factors influencing clinical practice. *Eur Heart J* 1998; **19**: 74–9.
46. Martinez M, Agustí A, Arnau JM, Vidal X, Laporte JR. Trends of prescribing patterns for the secondary prevention of myocardial infarction over a 13-year period. *Eur J Clin Pharmacol* 1998; **54**: 203–8.
47. Llop R, Ferrer A, Agustí A, Vidal X, Arnau JM, Laporte JR. From clinical trials to clinical practice; oral anticoagulation among patients with non-rheumatic atrial fibrillation. *Eur J Clin Pharmacol* 1997; **53**: 1–5.
48. Vallano A, Aguilera C, Arnau JM, Banos JE, Laporte JR. Management of post operative pain in abdominal surgery in Spain: a multicentre drug utilization study. *Br J Clin Pharmacol* 1999; **47**: 667–73.
49. Cochrane AL, Moore F. Expected and observed values for the prescription of vitamin B12 in England and Wales. *Br J Prev Soc Med* 1971; **3**: 147–51.
50. Shaywitz BA. Ritalin nation: rapid-fire culture and the transformation of human consciousness. *New Engl J Med* 1999; **340**: 1693.

7

A View From Industry

DALE B. GLASSER and GRETCHEN S. DIECK
Pfizer Inc, 235 East 42nd Street, New York, NY 10017, USA

INTRODUCTION

Innovative and successful pharmaceutical manufacturers invest heavily in research and development, in order to achieve increases in profit and growth through the launch of new products. The Pharmaceutical Research and Manufacturer's Association (PhRMA) in the US estimates that companies will invest 24 billion dollars in 1999 to discover and develop new drugs, which represents a 14% increase over 1998. In the 1990s more than 330 new medications have been approved by the US Food and Drug Administration (FDA), to treat conditions that affect millions of people.¹

Epidemiology has played an increasingly important role over the past 15 years in all phases of drug development, commercialization, and safety. Drug development has long relied on basic science research and on clinical studies of experimental rather than observational design. In the areas of safety and postmarketing surveillance, the pharmaceutical industry and regulatory authorities initially relied solely on clinical trials or on the spontaneous reporting system of the FDA or other regulatory authorities for information on safety. Epidemiology began its career in the pharmaceutical industry being used defensively, usually in response to legal or regulatory issues. However,

most large pharmaceutical companies now recognize the proactive contribution that may be made by epidemiology. In addition to its traditional role in drug safety and risk management, pharmacoepidemiology has become increasingly important in providing support to product planning, marketing, and market analysis.

PHARMACEUTICAL INDUSTRY GOALS AND OBJECTIVES

The primary goal of the pharmaceutical industry is to improve the public's health and well-being through the development and sale of safe and effective pharmaceutical products. Manufacturers devote major efforts to research and development, as well as to marketing those products that have been approved. Products must be sold at a profit for a manufacturer to remain economically viable. These profits, in turn, help to support research and development and lead to new drugs. Throughout the entire process of drug development and marketing, government regulatory authorities, such as the US FDA and similar agencies in other countries, regulate the industry.

Most drug regulatory authorities require that a pharmaceutical manufacturer demonstrates that

its product (either drug or device) is safe and effective before approving the product for sale. Beyond meeting regulatory needs, the manufacturer has an ethical obligation to monitor its products for safety on an ongoing basis, in order to develop and maintain product labeling that reflects current knowledge. The manufacturer's product label directly affects the safe prescribing of drugs by physicians and safe use of the product by the public.

In order to establish safety and efficacy, the manufacturer must have a basic understanding of the target patient population for both the successful development and sale of the product. To remain competitive, a manufacturer must be able to predict unmet medical need several decades into the future and understand the characteristics of the individuals afflicted with the various conditions of interest. Pharmaceutical research is extremely costly and time consuming. Given limited resources, decisions must be made as to which of many candidate drugs to develop. Once a drug is approved for sale, patient characteristics play an important role in developing marketing strategies and helping to maximize marketing resources by identifying specific target populations appropriate for the particular medication.

Throughout the lifespan of a drug, from discovery through the entire marketing and sales cycle, industry has an ongoing obligation to comply with worldwide regulatory requirements. Although few would seriously dispute the need for regulation, the pharmaceutical industry does need to function in a rational regulatory environment.² Implicit in this goal is the need for a logical basis for drug approval, the need for a rational and balanced approach to both pre- and postmarketing surveillance of drug safety, and the need for a regulatory environment relatively devoid of non-scientific pressures.²

Crucial to the continued success of the pharmaceutical industry in providing effective treatments for serious medical conditions is freedom from irresponsible attacks on safety issues. This is not to suggest that the industry should not be held accountable for its actions, but rather that challenges of drug safety should be motivated by consistently applied scientific principles.² To

minimize such attacks, there is need for rapid, efficient, and scientifically sound methods for evaluating drug safety. A proper understanding of the strengths and limitations of these methods is needed by government, special interest groups, the media, physicians, and the pharmaceutical industry itself.² Sales of a product, either drug or device, can be irreparably damaged or the product itself withdrawn from the market based on unwarranted and insupportable attacks, as was the case with Bendectin.³⁻⁵ Even when the safety of the product has subsequently been established, the public's trust may have eroded in both the product and the industry itself.

The manufacturer's responsibility is to develop safe and effective products, have the capability to rapidly assess the health needs and the safety of the population, and encourage and support the public's confidence in the actions of industry and government.² The methods of pharmacoepidemiology can contribute to each of these areas.

UTILITY OF PHARMACOEPIDEMOLOGY

Epidemiology makes its greatest contribution to the pharmaceutical industry in support of several functions: evaluating drug safety, new product planning and portfolio development, and commercialization or marketing of drugs. The observational methods used in these functions are often described by other terminology, such as cost-effectiveness research (see Chapter 35), outcomes research (see Chapter 36), or quality of life research (see Chapter 36). Regardless of nomenclature, pharmacoepidemiology is playing an expanding role within the pharmaceutical industry.

EVALUATING DRUG SAFETY

Evaluating drug safety is the primary use of epidemiology within the pharmaceutical industry. Drug manufacturers have traditionally relied on two major sources for information on the safety of drugs: the clinical trials supporting the New Drug Application (NDA) and, once the drug is marketed, spontaneous reports received throughout

the world. Both are useful and have a place in assessing drug safety, but both have limitations that can be addressed, in part, by the proper use of formal observational epidemiology. Epidemiologic studies can complement these two sources of data to provide a more comprehensive and pragmatic picture of the safety profile of a drug.

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 relatively 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 (occasionally as high as 7000). Using the “rule of three”,^{6,7} 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 (see also Chapter 3). 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.^{7,8} Thus, clinical trials are usually only large enough to detect events that occur relatively frequently, and are generally inadequate to address all potential safety issues related to a particular drug.⁹

An additional limitation of clinical trials with respect to drug safety is the strict inclusion/exclusion criteria common in clinical trials. 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 may be excluded from some trials.^{10–13} 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 aborted or reversed with proper treatment. In contrast, once a drug is marketed, it is used by patients in “real-life”

situations, who may have a constellation of disorders for which they are being treated simultaneously. Patients may also be taking over the counter medications, “natural remedies,” or illicit drugs unbeknownst to the prescribing physician. The interactions of various drugs and treatments may result in a particular drug having a different safety profile in a postmarketing setting compared to the controlled premarketing environment.¹⁴ A recent example is the drug Posicor (mibefradil), which was voluntarily withdrawn from the market after less than a year by the manufacturer as a result of new information about multiple potentially serious drug interactions.¹⁵ Compliance in taking the drug as prescribed may also differ between closely monitored trials and general postapproval 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 also Chapters 10 and 11).^{16,17} Signals are used to generate hypotheses, which then may be studied in observational or interventional studies. Spontaneous reports must be interpreted within the context of the strengths and limitations of the system.¹⁷ These reports are subject to many biases, which are unmeasured and unmeasurable. Events are generally underreported and the decision of which events to report is potentially strongly affected by bias.¹⁷ The effects of these biases differ among drugs and differ 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.^{16–19} Certain types of event seem to be more likely to be reported, such as those which are serious and/or unlabeled,^{20,21} events that occur rarely in the general population,²² those that occur acutely with drug administration,²² and those associated with publicity in the lay or professional media.¹⁶ The frequency of reporting varies by drug class and

drug company.²¹ The number of reports is not equal to the number of patients, since events are often reported several times.²³ Valid incidence rates cannot be generated from spontaneous reporting systems, since neither the true numerator nor true denominator is known,¹⁷ and reporting rates between drugs cannot be validly compared; therefore, relative safety cannot be assessed.²³ Additionally, the events reported have an underlying background rate in the population, even in the absence of drug treatment.¹⁷

Notwithstanding these important limitations, the spontaneous reporting system has been successfully used in a number of circumstances to alert a manufacturer to a potentially high frequency of serious adverse events in a newly launched drug. One example is temafloxacin (Omniflox), which was approved by the FDA in January 1992. By June 1992, the drug was voluntarily withdrawn from the market by the manufacturer, following reports of six deaths and more than 70 other serious adverse events, including hemolytic anemia, renal failure, severe hypoglycemia, and anaphylaxis. In the first four months after marketing, an estimated 174 000 individuals took this drug,²⁴ allowing the rapid observation of serious adverse events less frequent than those observed in clinical trials. Another example is that of mibefradil (Posicor), a calcium channel blocker first marketed in the US in August 1997. Three drugs (astemizole, cisapride, and terfenadine) were listed as having an interaction with mibefradil at the time of approval. Through spontaneous reports (as well as continued clinical studies), more than 25 drugs were identified that were potentially harmful if used with mibefradil. Mibefradil had no known special benefits that could not be met with other drugs. The manufacturer and the FDA decided that the number and diversity of drugs with which it interacted could not be practically handled via the usual warnings in the label, and the risk–benefit profile was unfavorable. The drug was voluntarily withdrawn from the US market in June 1998.¹⁵

Of great concern is the misinterpretation and dissemination of misinformation based upon reports received through the spontaneous reporting system. A well publicized example deals with

the claims that the antidepressant fluoxetine (Prozac) might be associated with suicidal behavior and violent aggression towards others.²⁵ Intense publicity contributed to loss of market share for this product although practitioners and an FDA Advisory Committee had taken the position that suicidal behavior is an inherent part of depression and not associated with toxic effects of the drug, and that there is no scientific evidence linking aggressive behavior with this agent.^{26,27} The FDA Advisory Committee did suggest that the manufacturer further support the evidence of the drug's safety with prospective epidemiologic studies. Subsequent studies and re-analyses of clinical trial data did not support the speculation that fluoxetine increased the risk of suicide or aggressive behavior, and millions of patients worldwide have benefited from this therapy.^{28–31}

In order to evaluate safety signals arising from spontaneous reporting systems, one needs to know as much as possible about the population using the drug. For example, knowledge about the distribution of age, gender, concomitant illnesses, and medications in users of a particular drug can provide information necessary to estimate the expected background rates of events that one might observe. A number of commercial vendors such as NDC Health Information Services, HCIA Inc., and IMS Health provide extensive information about the use and sales of prescription products (see also Chapters 24 and 29). Although information about actual drug ingestion is not available, one can make some assumptions about the frequency of use from calculating the interval between refills in longitudinal resources. Additionally, information about the frequency of off-label use or the frequency of coprescribing with medications that are contraindicated may be explored.

There has been concern for many years that the FDA was taking too long to review New Drug Applications (NDAs) and make decisions on approval. While a thorough review of the safety and efficacy data is necessary, unnecessary delays due to staff shortages at the FDA delayed many valuable, and in some cases lifesaving, medications from being available to the large numbers of patients who could benefit. The community of

AIDS–HIV activists was instrumental in forcing change. In 1992, the Prescription Drug User Fee Act (PDUFA) was passed. This enabled the FDA to hire hundreds of additional reviewers with funds provided by the sponsors (manufacturers) submitting an NDA. In return, manufacturers could expect a decision on approval within one year of submitting an NDA (6 months if the drug was approved for expedited review by the FDA). In 1992, 26 new drugs were approved by the FDA, compared to 53 in 1996. The act expired in 1997 and was subsequently renewed by Congress. There has been increased emphasis on postmarketing surveillance since that time, and in some cases companies have been asked to perform formal observational studies as part of the approval commitment.

Formal Epidemiology Studies

There are many safety issues relevant to the pharmaceutical industry 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. For example, case reports of a few patients with primary pulmonary hypertension exposed to appetite suppressant drugs led to a formal epidemiologic study documenting this association and strengthened labeling for the drug.³²

Epidemiologic studies may also be the source of hypotheses or signals that later are found not to be supported. For example, Pahor and associates suggested that calcium antagonists were associated with gastrointestinal hemorrhage³³ and cancer.^{34,35} Later studies did not confirm these findings, but sales of the products had been compromised.^{36–41}

Epidemiologic studies may also be used to examine the comparative risks associated with particular drugs within a therapeutic class. For example, one large study determined that, among anti-ulcer drugs, cimetidine was associated with the highest risk of developing symptomatic acute liver disease.⁴² Other studies examined the risk of hip fractures in users of benzodiazepines and

found that users of long acting agents were at greater risk than those using short acting agents.^{43,44}

Epidemiologic studies are extremely useful to place the incidence of adverse events observed during 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 rates for individuals with comparably severe disease.^{45–47}

Epidemiologic methods have also been used extensively to examine possible teratologic effects of various agents (see also Chapter 42). 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. Experimental methods such as randomized, double blind clinical trials are not feasible to address these issues. Instead, epidemiologic methods are necessary to gather and evaluate the information in the population actually using the drug. Examples of epidemiologic studies examining possible teratogenicity include those studying diazepam use and oral clefts,⁴⁸ spermicide use and Down's syndrome, hypospadias, and limb reduction deformities,⁴⁹ and Bendectin use and oral clefts, cardiac defects, and pyloric stenosis.^{50,51} In certain circumstances manufacturers may choose to set up pregnancy registries to obtain information about the use of their products.^{52,53}

Epidemiologic methods have also been used to study cancers in individuals exposed to drugs *in utero*, periconceptually, or immediately after birth. A classic example is the association between maternal use of DES and clear-cell adenocarcinoma of the vagina.^{54,55} Other examples include the possible association between prenatal exposure to metronidazole and childhood cancer,⁵⁶ and childhood cancers and the use of sedatives during pregnancy.⁵⁷ A number of studies have examined the potential association between childhood cancer and exposure to vitamin K in the neonatal period.^{58–61}

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).^{62–66} 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.

Modern chemotherapy for childhood cancer has only been in use since approximately 1970 and it is only fairly recently that large numbers of children are being cured of cancer, allowing for the estimation of the long-term risks associated with the use of cytotoxic agents. Epidemiologic studies have documented the association between iatrogenic leukemia and treatment with alkylating agents or epipodophyllotoxins for previous cancers.^{67,68} Second malignant neoplasms (of the solid tumor type) have also been associated with the use of alkylating agents and antitumor antibiotics.^{69,70} 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.^{71–73} Decreased bone mineral density has also been documented following chemotherapeutic treatment in childhood.^{74,75}

Survivors of adult cancers have also been the subjects of studies examining associations between chemotherapy and late effects. Examples include findings of decreased bone mineral density in women treated with cytotoxic agents for breast cancer⁷⁶ and in men and women treated for Hodgkin's disease,^{77,78} which may be due to a direct effect of treatment on bone, a secondary effect mediated via gonadal toxicity, or a combination of the two.

MEDICAL DEVICES

Some pharmaceutical manufacturers also produce medical devices. Epidemiologic methods may also be used to evaluate the safety of these devices, including those that are permanently surgically implanted, such as breast augmentation devices or prosthetic joint replacements (see also Chapter 41).

Premarketing testing of devices must, by definition, differ from that of drugs. Surgical implants are usually designed for long term or life-long use, and clinical trials are not feasible to study effects with long latency periods. Surgical techniques also differ widely among surgeons, and may influence later adverse events. Placebo comparisons are not possible to study surgical implants for obvious reasons. It may be only when the device has been in widespread use for many years that some adverse effects are noticed. At this point, epidemiologic studies may be useful.

The regulation of medical devices by the FDA also differs substantially from that of drugs, and will only be briefly discussed here (see also Chapter 41). Prior to the Medical Device Amendments of 1976 to the Federal Food, Drug and Cosmetic Act, devices could be marketed without review by the FDA.⁷⁹ Now, devices are divided into three classes (I, II, and III) of increasing theoretical potential hazard, and consequently increasing regulatory rigor. Medical devices are also classified by whether they are “pre- or post-amendment” (marketed prior to or after the amendment of 1976), “substantially equivalent” to a pre-amendment device, and additionally, whether they are “implanted” (to remain for more than 30 days), “custom”, “investigational”, or “transitional” (previously defined as drugs, but now defined as devices, i.e., injectable silicone and intraocular lenses). If a manufacturer can provide data establishing “substantial equivalence” to a pre-amendment device, a premarketing notification to the FDA (“510[k]”) is all that is required, while full testing and approval are required for all postamendment devices that are not substantially equivalent. There has been considerable controversy over the ability to “grandfather” new devices through the 510(k) procedure by showing substantial equivalence to pre-amendment devices that had never been formally approved through the PMA process (Pre-market Approval Application, similar to an NDA).

The best known recent example of an investigation regarding the safety of a medical device deals with silicone breast implants. In 1992, the Commissioner of the FDA recommended a temporary moratorium on the use of such devices. Lawsuits

were brought by hundreds of thousands of women against the manufacturers of the implants claiming that the silicone had caused a variety of connective tissue disease. Over 7 billion dollars in judgments and settlements were paid by the companies.⁸⁰ A panel of experts appointed by the Institute of Medicine concluded that women with breast implants did not have an increased risk of cancer, immunologic diseases, or neurological problems over women without such implants.⁸¹ They did conclude that implants often led to complications requiring surgery to correct them.

EPIDEMIOLOGY IN SUPPORT OF THE DEVELOPMENT AND MARKETING OF PHARMACEUTICALS

New Product Planning and Portfolio Development

The future success of a pharmaceutical company lies in the value of its product pipeline. Product planning (the development of drugs) is a critical function within innovative pharmaceutical companies, because of the continuing need for new and promising developmental drugs to fill their product pipelines. Pharmacoepidemiology has become an effective tool in this area. Basic epidemiologic techniques have been useful for defining markets, for determining how a drug is actually being used in the population, and for determining public health needs, especially where no true commercial market already exists. Pharmacoepidemiology is also well suited to identifying high-risk groups such as the elderly, the poor, expectant mothers, and so on. Such knowledge aids a drug company in positioning its drugs in the market place and in weighing risks and benefits of therapy.²

Estimating the incidence and prevalence of a disease is crucial for estimating the potential market value of a drug, allowing a company to allocate limited research funds appropriately. This is especially important, given that drugs often take up to 15 years to develop and cost an average of \$500 million to develop from the laboratory to use in humans in the US.^{82, 83} Only five out of every 5000 products that enter preclinical testing eventually make it to human testing, and only one of

those five products is approved for sale.⁸⁴ Successful companies carefully choose which early candidates in their pipeline to progress. Information regarding the descriptive epidemiology of a condition may lead to decisions to progress a candidate drug on a “fast track.”

This type of information may be available through 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. Alternatively, this information may be derived from population based studies commissioned by industry, although the cost and time investment is considerably higher.

Epidemiologic studies can help a pharmaceutical company describe the prevalence of a condition in which they have an interest,^{85, 86} describe the natural history of a condition prior to the introduction of their drug,⁸⁷⁻⁹⁰ and describe the frequency with which complications of that condition occur.⁹¹ The examples in the preceding sentence all relate to the Olmsted County Study of Urinary Symptoms and Health Status Among Men, funded by Merck Laboratories in support of the development of finasteride (Proscar). In addition to prevalence, epidemiologic studies can estimate the burden (cost and disability) associated with specific conditions, as has been done with migraine,^{92, 93} asthma,⁹⁴ and Alzheimer’s disease.⁹⁵ The burden of disease estimates the cost, both human and economic, to sufferers and to society and is helpful for valuing a drug in terms of unmet medical need.

These same epidemiologic methods may be used for “evergreening” purposes (exploring new indications for existing products). Additionally, products with multiple indications may warrant investigations into the joint prevalence of specific comorbid conditions.

Marketing Support

Once a drug is approved, the manufacturer is interested in monitoring its use and the type of patient who receive the drug. Observational studies may be informative as to the frequency of

off-label use⁹⁶ and use of multiple medications.^{97,98} 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, or changes in the characteristics of users over time.^{99–101} Epidemiologic methods may also be useful to study medication use among high risk populations, such as the elderly with cognitive impairment^{102,103} or pregnant women.¹⁰⁴

Epidemiologic methods may be used to benefit marketed compounds in several ways. Product positioning efforts may be informed by epidemiologic studies suggesting either special risk or special benefits in certain subpopulations. Disease awareness activities (professional and patient educational programs) may also be useful, particularly when the indication for a drug is a symptom or condition that the patient himself must bring to the doctor's attention, such as erectile dysfunction. Information on patterns of health-care utilization may also help target marketing efforts.

The FDA's Division of Drug Marketing Advertising and Communication (DDMAC) monitors advertising of drugs. Any claim in an advertisement must be based on scientifically sound studies. An example of a recent drug advertisement that used observational data is a professional journal advertisement for norethindrone acetate and ethinyl estradiol tablets (Estrstep), which cited a reference by Rosenberg and associates.¹⁰⁵

BENEFICIAL EFFECTS OF DRUGS

Epidemiologic methods are also useful to quantify the beneficial effects of drugs (see also Chapter 34). Study endpoints may vary from outcomes such as well-being or quality of life to more quantitative variables such as blood pressure level, direct and/or indirect cost savings, and utilization of the health care system. Examples include studies to document improvements in health related quality of life associated with drugs for migraine, benign prostatic hyperplasia, allergic rhinitis, osteoarthritis, and rheumatoid arthritis.^{106–111}

Manufacturers are also using the tools of pharmacoepidemiology to carry out economic

studies (see also Chapter 35). Studies that demonstrate an economic advantage are useful for marketing if a manufacturer can demonstrate that use of its product is equally effective as but less costly than a competitor's. These studies may be used to justify inclusion of brand name products on formularies of health maintenance organizations (HMOs), hospitals, and state Medicaid programs. Recent studies include the economic advantages of the addition of selective serotonin reuptake inhibitors to speed the improvement of depression,¹¹² the cost-effectiveness of several agents for hypertension,¹¹³ the investigation of cost-effectiveness of a treatment for mild to moderate Alzheimer's disease,¹¹⁴ the measurement of direct and indirect costs of treating allergic rhinitis,¹¹⁵ and the quality of life and health economic benefits, including work productivity, associated with improved glycemic control.¹¹⁶

Beneficial effects of drugs with respect to prevention of disease have been studied using epidemiologic techniques. Examples include the decrease in risk of cardiovascular disease, a protective effect against worsening of osteoarthritis of the knee, the protection against tooth loss associated with post-menopausal estrogen replacement therapy,^{117–119} and the possible decreased risk of cancer of the digestive tract and certain other sites in users of aspirin or nonsteroidal anti-inflammatory agents.^{120–123}

ISSUES IN PHARMACOEPIDEMOLOGY

RESOURCES FOR PHARMACOEPIDEMOLOGY

In order to respond rapidly and responsibly to safety issues, industry must have access to high quality, valid data resources. As a result of this need, an area of pharmacoepidemiology that has experienced considerable growth is the development and use of record linkage and automated databases.¹²⁴ Several databases will be discussed in depth in other chapters of this book (see Chapters 15–25) and the matter of databases will only be discussed generally here. Existing 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 relatively 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 both new and existing epidemiologic databases that contain information on drug usage and health related outcomes;¹²⁵⁻¹²⁸ however, further improvement is needed. A variety of different data sources are necessary since no single data system is sufficient to meet all needs. Existing automated databases have some major limitations. Often a study is needed regarding a newly marketed product. Sufficient numbers of users may not yet be recorded in the databases, or the product may not have yet been marketed in the particular country included in the database. Some data resources suffer from a considerable "lag-time" between data entry and availability for research purposes. Many health maintenance organization databases have overall enrollments of hundreds of thousands of members, but these numbers may be inadequate to study the risks of extremely rare events associated with a specific drug.

Many databases were designed for administrative purposes, rather than for epidemiologic studies. As a result, specific information needed to assess a safety issue may be unavailable and quality of medical information may be inadequate. Continuing studies of the validity of this information are crucial.¹²⁹⁻¹³³ The usefulness of certain databases to answer a specific question can be severely limited if the data are not properly validated.

Even established resources of high caliber are not guaranteed to continue. Quality data sources are a function of the financial and administrative

support received. Changes in support, such as loss of funding or alteration in study administration, may drastically affect the format of the data or the accessibility of the resource. An example is the Saskatchewan Drug database, which was substantially altered for a period of time because of administrative changes in the billing system (see Chapter 20). During the period from middle of 1987 until the end of 1998, the codes identifying individuals receiving particular drugs were replaced by family codes only. Outcome variables could no longer be linked to individuals who had received particular drugs. The Saskatchewan database previously been a rich source of data for studies such as a case-control study of cyclic antidepressants and the risk of hip fracture¹³⁴ and a cohort study of NSAIDs and serious gastrointestinal events.¹³⁵ However, with these changes in the billing system, the database would no longer be able to support such studies. After discussions with epidemiologists, the Saskatchewan health authorities reinstated the previous system of individually identifying patients,^{136,137} and the database is once again being used for research purposes.^{126,138,139}

As another example, The General Practice Research Database (formerly known as VAMP) in the United Kingdom (see Chapter 23) suffered from uncertain funding for a number of years, and funds were provided primarily through two principal licensees of the data. In April 1999, the Department of Health announced that the Medicines Control Agency (MCA) would be responsible for the management of the GPRD and confirmed that 3 million pounds sterling would be invested in the database over the next five years to further develop the database, improve data submission procedures, improve feedback to the practitioners who contribute the data, and enhance services provided to the users of the database.

MISUSE OR MISINTERPRETATION OF PHARMACOEPIDEMIOLOGIC STUDIES

Careful, scientifically sound research carried out using quality resources does not guarantee that a study's findings will be appropriately used. Assuming that a safety issue has been suitably

addressed using appropriate data resources, the study findings may be improperly interpreted or misused. The irresponsible use or misinterpretation of pharmacoepidemiology studies may have direct implications on the industry because of tremendous costs due to resulting liability or to the drug being removed from the market, and implications to the patients who could benefit from the medications. Regulatory agencies are affected by having to devote scarce resources to evaluating erroneous safety issues in order to make regulatory decisions. Ultimately, a disservice has been carried out to the public by generating unwarranted fears, by the removal of safe and effective drugs, and by higher costs for pharmaceuticals.

A well known example of study results being inappropriately interpreted is the controversy over the purported relationship between spermicide use and birth defects. Early studies suggested an association between spermicide use and a number of different birth defects,^{140,141} but subsequent studies did not confirm these findings.¹⁴²⁻¹⁴⁷ Despite the acknowledged weaknesses of the study design suggesting an association,^{148,149} the consistent and overwhelming evidence supporting the safety of the contraceptive method, as well as the opinion of the FDA and its Fertility and Maternal Health Drugs Advisory Committee,¹⁵⁰ a court decision was found in favor of a plaintiff claiming that spermicide use caused their child's birth defects.¹⁵¹

The unfavorable consequences of such a finding to the individual drug manufacturer are potentially enormous. An immediate result is the loss of the lawsuit with its financial implications. Additional cases are likely and the manufacturer is faced with years of pending litigation. Less tangible, but perhaps more important to the industry, is that the confirmation of safety brought forward by pharmacoepidemiologic research in the case of spermicide use and birth defects was trivialized in light of the findings made by the court. A certain amount of control can be exercised by drug companies in insuring that epidemiologic studies are thoughtfully designed and carefully executed, but industry is almost powerless to impact on how the studies are interpreted and used.

Misuse of epidemiologic results by the media may also result in a useful drug being precipitously withdrawn from the market. Bendectin, used for nausea during pregnancy, was marketed in the US from 1956 through 1983. The manufacturer voluntarily withdrew the drug from the market in 1983 because of the cost of defending the large number of product liability lawsuits filed following extensive publicity suggesting that the drug was teratogenic. Numerous epidemiologic studies in varying settings with various designs were performed to examine this issue. The results did not support the suspicion that Bendectin was a teratogen.^{4,5} Because of its withdrawal, this drug is no longer available to the 10-25% of all pregnant women who would potentially benefit from its use.¹⁵² Neutel *et al.* reported that in the year following withdrawal of Bendectin from the market, hospitalizations for vomiting in pregnancy (per thousand live births) in Canada rose by 37% and by 50% the year after that, with similar findings in the US. Their estimates of excess hospital costs over the years 1983-1987 was \$16 million in Canada and \$73 million in the US.¹⁵³

Misinterpretation of epidemiologic studies in the media perpetuates the impression that the discipline is fraught with weaknesses through generating controversy over study results. 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 to be of questionable usefulness. The dissemination of misinformation promotes needless anxiety on the part of both patients and physicians. Illustrations of this concern are the unnecessary abortions that most likely resulted from an irresponsibly written newspaper article about the teratogenicity of Bendectin.³

BALANCING LOGISTICS AND UTILITY

Epidemiologists in the pharmaceutical industry must balance the high cost and extended time-frame necessary for large prospective studies with usefulness to the company. When there is a potential safety concern, information is needed quickly in order to evaluate the balance between

benefit and risk. In other cases of less immediate concern, by the time study results become available, the drug may be off patent. It is important to quantify the risk associated with drug products and to do it in a rapid and efficient way.

FUTURE DIRECTION OF PHARMACOEPIDEMIOLOGY

QUANTITY OF RESOURCES

Improving the number and variety of pharmacoepidemiologic data resources and methodologies is necessary to insure that drug manufacturers can validly and rapidly address safety questions using appropriate study populations and research designs, in both routine as well as acute drug safety situations. The pharmaceutical industry can play an active role in the enrichment of data sources by supporting existing resources, in addition to helping to develop new databases and new postmarketing surveillance methods. The active support of research groups and existing databases, even in the absence of a research question, will serve two purposes. First, it will help ensure that quality resources continue even though major funding sources disappear. Also, additional funds above and beyond administrative costs may be directed toward improving the database and toward making it more flexible or more appropriate for use by the drug manufacturer.

Industry support is critical to the development of new database resources. Research groups may have access to various types of information, but lack the financial resources to develop the information into a usable database. Pharmaceutical companies actively seeking new data resources should not overlook any potentially valuable sources of information. Industry guidance and funding in such cases may be critical to the development of a viable data resource. Data linkage between existing databases is another area where industry support can directly promote the growth of resources by providing financial support.

The development of new postmarketing surveillance methods is an additional area that should

continue to grow. New methods are needed to accrue large numbers of individuals on particular therapeutic regimens rapidly and to be able to follow them prospectively for both beneficial and adverse health outcomes. Exploring new methods may be costly and demands a special commitment from drug companies that recognize that such an investment may ultimately pay off by providing additional means to evaluate drug safety and effectiveness. Industry epidemiologists should also keep abreast of new study designs that may offer improved methods of evaluating the risks and safety associated with pharmaceutical products. For example, a recent publication used the case-crossover design to investigate the association between road traffic accidents and benzodiazepine use.¹⁵⁴

Finally, there are a limited number of formal training programs in pharmacoepidemiology at universities. This discipline is not supported as strongly by manufacturers as are clinical or basic science fellowships. In order to meet the increasing needs of industry with respect to trained epidemiologists, drug manufacturers should increase their role in supporting training and fellowship programs. Such support not only helps insure sufficient epidemiologic manpower requirements with specific training in drug safety, but also provides a structure for the implementation of high quality pharmacoepidemiologic research.

QUALITY OF RESOURCES

If pharmacoepidemiology is to continue to play an important role in the evaluation of the safety of drugs, then the resources and the research itself must be of the highest quality. One means of improving the quality of research and resources is by setting standards for investigations. High standards for study design and execution should be encouraged. Industry can exercise some control by supporting investigators with proven track records and by refusing to provide funding for protocols of poor quality. Industry can help improve the quality of various data resources by sponsoring seed money to set up such resources, and pilot studies to validate computerized information with the corresponding medical records.

Ultimately, poorly carried out studies or studies using data of questionable quality are damaging to both the industry and the credibility of the field of epidemiology.

PROMOTING A GREATER UNDERSTANDING OF PHARMACOEPIDEMOLOGY

Epidemiologic studies are of limited use to the pharmaceutical industry if they are not well understood. Greater understanding of the strengths and limitations of epidemiology is needed by the public, the media, the government, and by industry itself.² These diverse groups have common interests and, through their joint efforts, the discipline of epidemiology may be improved by focusing support, assessing study quality, and advancing greater understanding of the field.² Epidemiologists are also responsible for education about the discipline within their own companies so that the results of observational studies may be best leveraged according to the needs of the company.

COMBINING EFFORTS WITHIN THE INDUSTRY

Within the pharmaceutical industry there is a collegial relationship among epidemiologists and drug safety specialists. A number of venues provide the opportunity for industry representatives to work together, such as the Pharmaceutical Research and Manufacturing Association (PhRMA), and to work with members of regulatory authorities worldwide on issues of mutual interest (Council for International Organizations of Medical Sciences (CIOMS)). Both industry and regulatory authority professionals collaborate with academic researchers through the International Society for Pharmacoepidemiology (ISPE).

Combining efforts among drug manufacturers on matters of mutual interest benefits the individual companies through shared knowledge regarding data resources and evolving epidemiological methods. Pooled funding also benefits the industry as a whole through the maintenance and development of improved resources. The field could also

benefit from joint projects carried out by two or more pharmaceutical companies who want to explore a particular issue.

CONCLUSIONS

Epidemiology can make a significant contribution to the development and marketing of safe and effective pharmaceutical products worldwide. It may also serve to facilitate the regulatory process and provide a rational basis for drug safety surveillance. Like any other tool, it must be properly understood and utilized. Industry has a continuing opportunity to influence the development of the field and the responsibility to do so in a manner that will not only expand resources but will assure scientific validity. Achieving this goal requires financial and intellectual support as well as also requiring a better understanding of the nature of the discipline and its uses. The growing number of epidemiologists within the industry is a positive sign, but it is also clear that there is still need for greater commitment. A greater appreciation of the role of epidemiology is needed within the industry, the medical community, regulatory agencies, and the media.

REFERENCES

1. Holmer AF. Drug companies add 39 new treatments to the nation's medicine chest, increase R & D investment to \$24 billion. In: New drug approvals in 1998. Washington, DC: PhRMA, January 1999; 1 (http://www.phrma.org/chart/archive/1998/nda_c98.html).
2. Sachs RM, Dieck GS. Industry perspectives and the contributions of pharmacoepidemiology to public health. In: Edlavitch S, ed., *Pharmacoepidemiology, Proceedings of the Third International Conference on Pharmacoepidemiology*. Chelsea, MI: Lewis, 1989; 43–49.
3. Brent RR. The Bendectin saga: another American tragedy. *Teratology* 1983; **27**: 283–6.
4. Brent RL. Bendectin: review of the medical literature of a comprehensively studied human nonteratogen and the most prevalent tortogen-litigen. *Reprod Toxicol* 1995; **9**: 337–49.
5. McKeigue PM, Lamm SH, Linn S, Kutcher JS. Bendectin and birth defects: I. A meta-analysis

- of the epidemiologic studies. *Teratology* 1994; **50**: 27–37.
6. Strom BL. The promise of pharmacoepidemiology. *Annu Rev Pharmacol Toxicol* 1987; **27**: 71–86.
 7. Sackett DL, Haynes RB, Gent M, *et al.* Compliance. In: Inman WHW, ed., *Monitoring for Drug Safety*, 2nd edn. Hingham: MTP Press, 1986; 427–38.
 8. Congress of the United States, Office of Technology Assessment. Kemp KB, ed. *Postmarketing Surveillance of Prescription Drugs*, document No. OTA-H-189. Washington, DC: US Government Printing Office, 1982.
 9. Faich GA, Lawson DH, Tilson HH, *et al.* Clinical trials are not enough: drug development and pharmacoepidemiology. *J Clin Res Drug Dev* 1987; **1**: 75–8.
 10. Lasagna L. Should new drugs be evaluated in children? In: Lasagna L, ed., *Controversies in Therapeutics*. Philadelphia, PA: Saunders, 1980; 455–463.
 11. Yastrubetskaya O, Chiu E, O'Connell S. Is good clinical research practice for clinical trials good clinical practice? *Int J Geriatr Psychiatry* 1997; **12**: 227–31.
 12. Merton V. The exclusion of pregnant, pregnable, and once-pregnable people (a.k.a. women) from biomedical research. *Am J Law Med* 1993; **19**: 369–451.
 13. Gurwitz JH, Col NF, Avorn J. The exclusion of the elderly and women from clinical trials in acute myocardial infarction. *J Am Med Assoc* 1992; **268**: 1417–22.
 14. Rogers AS. Adverse drug events: identification and attribution. *Drug Intell Clin Pharm* 1987; **21**: 915–20.
 15. Food and Drug Administration. *Roche Laboratories Announces Withdrawal of Posicor from the Market*, FDA Talk Paper. 1998.
 16. Baum C, Kweder SL, Anello C. The spontaneous reporting system in the United States. In: Strom B, ed., *Pharmacoepidemiology*, 2nd edn. New York: Wiley, 1994; 125–37.
 17. Goldman SA. Limitations and strengths of spontaneous reports data. *Clin Ther* 1998; **20**: C40–4.
 18. Faich GA. International drug surveillance. *Drug Info J* 1985; **19**: 227–9.
 19. Sachs RM, Bortnichak EA. An evaluation of spontaneous adverse drug reaction monitoring systems. *Am J Med* 1986; **81**: 49–55.
 20. Moride Y, Haramburu F, Requejo AA, Begaud B. Under-reporting of adverse drug reactions in general practice. *Br J Clin Pharmacol* 1997; **43**: 177–81.
 21. Alvarez-Requejo A, Carvajal A, Begaud B, *et al.* Under-reporting of adverse drug reactions. Estimate based on a spontaneous reporting scheme and a sentinel system. *Eur J Clin Pharmacol* 1998; **54**: 483–8.
 22. Brewer T, Colditz GA. Postmarketing surveillance and adverse drug reactions: current perspectives and future needs. *J Am Med Assoc* 1999; **281**: 824–9.
 23. FDA, Division of Pharmacovigilance and Epidemiology. *Food and Drug Administration Adverse Drug Reaction System: Brief Description with Limitations of System*. 1997.
 24. Finch RG. The withdrawal of Temafloxacin: are there implications for other quinolones? *Drug Safety* 1993; **8**: 9–11.
 25. Teicher MH, Glod C, Cole J. Emergence of intense suicidal preoccupation during fluoxetine treatment. *Am J Psychiatry* 1990; **147**: 207–10.
 26. Ahmad SR. USA: Fluoxetine “not linked to suicide”. *Lancet* 1991; **338**: 875–6.
 27. FDA Psychopharmacologic Drugs Advisory Committee Meeting. 21 September 1991.
 28. Heiligenstein JH, Beasley CM Jr, Potvin JH. Fluoxetine not associated with increased aggression in controlled clinical trials. *Int Clin Psychopharmacol* 1993; **8**: 277–80.
 29. Coccaro EF, Kavoussi RJ. Fluoxetine and impulsive aggressive behavior in personality-disordered subjects. *Arch Gen Psychiatry* 1997; **54**: 1081–8.
 30. Warshaw MG, Keller MB. The relationship between fluoxetine use and suicidal behavior in 654 subjects with anxiety disorders. *J Clin Psychiatry* 1996; **57**: 158–66.
 31. Leon AC, Keller MB, Warshaw MG, Mueller TI, Solomon DA, Coryell W, Endicott J. Prospective study of fluoxetine treatment and suicidal behavior in affectively ill subjects. *Am J Psychiatry* 1999; **156**: 195–201.
 32. Abenham L, Moride Y, Brenot F, Rich S, Benichou J, Kurz X, Higenbottam T, Oakley C, Wouters E, Aubier M, Simonneau G, Begaud B. Appetite-suppressant drugs and the risk of primary pulmonary hypertension. International Primary Pulmonary Hypertension Study Group. *New Engl J Med* 1996; **335**: 609–16.
 33. Pahor M, Guralnik JM, Furberg CD, Carbonin P, Havlik R. Risk of gastrointestinal haemorrhage with calcium antagonists in hypertensive persons over 67 years old. *Lancet* 1996; **347**: 1061–5.
 34. Pahor M, Guralnik JM, Salive ME, Corti MC, Carbonin P, Havlik RJ. Do calcium channel blockers increase the risk of cancer? *Am J Hypertens* 1996; **9**: 695–9.
 35. Pahor M, Guralnik JM, Ferrucci L, Corti MC, Salive ME, Cerhan JR, Wallace RB, Havlik RJ. Calcium-channel blockade and incidence of cancer in aged populations. *Lancet* 1996; **348**: 493–7.
 36. Smalley WE, Ray WA, Daugherty JR, Griffin MR. No association between calcium channel

- blocker use and confirmed bleeding peptic ulcer disease. *Am J Epidemiol* 1998; **148**: 350–4.
37. Kelly JP, Laszlo A, Kaufman DW, Sundstrom A, Shapiro S. Major upper gastrointestinal bleeding and the use of calcium channel blockers. *Lancet* 1999; **353**: 559.
 38. Suissa S, Bourgault C, Barkun A, Sheehy O, Ernst P. Antihypertensive drugs and the risk of gastrointestinal bleeding. *Am J Med* 1998; **105**: 230–5.
 39. Jick H, Jick S, Derby LE, Vasilakis C, Myers MW, Meier CR. Calcium-channel blockers and risk of cancer. *Lancet* 1997; **349**: 525–8.
 40. Rosenberg L, Rao RS, Palmer JR, Strom BL, Stolley PD, Zaubler AG, Warshauer ME, Shapiro S. Calcium channel blockers and the risk of cancer. *J Am Med Assoc* 1998; **279**: 1000–4.
 41. Vezina RM, Lesko SM, Rosenberg L, Shapiro S. Calcium channel blocker use and the risk of prostate cancer. *Am J Hypertens* 1998; **11**: 1420–5.
 42. Garcia Rodriguez LA, Wallander MA, Stricker BH. The risk of acute liver injury associated with cimetidine and other acid-suppressing anti-ulcer drugs. *Br J Clin Pharmacol* 1997; **43**: 183–8.
 43. Ray WA, Griffin MR, Downey W. Benzodiazepines of long and short elimination half-life and the risk of hip fracture. *J Am Med Assoc* 1989; **262**: 3303–7.
 44. Cummings SR, Nevitt MC, Browner WS, Stone K, Fox KM, Ensrud KE, Cauley J, Black D, Vogt TM. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. *New Engl J Med* 1995; **332**: 767–73.
 45. Tennis P, Cole TB, Annegers JF, Leestma JE, McNutt M, Rajput A. Cohort study of incidence of sudden unexplained death in persons with seizure disorder treated with antiepileptic drugs in Saskatchewan, Canada. *Epilepsia* 1995; **36**: 29–36.
 46. Derby LE, Tennis P, Jick H. Sudden unexplained death among subjects with refractory epilepsy. *Epilepsia* 1996; **37**: 931–5.
 47. Leestma JE, Annegers JF, Brodie MJ, Brown S, Schraeder P, Siscovick D, Wannamaker BB, Tennis PS, Cierpial MA, Earl NL. Sudden unexplained death in epilepsy: observations from a large clinical development program. *Epilepsia* 1997; **38**: 47–55.
 48. Rosenberg L, Mitchell AA, Parsells JL, *et al*: Lack of relation of oral clefts to diazepam use during pregnancy. *New Engl J Med* 1983; **309**: 1282–5.
 49. Louik C, Mitchell AA, Werler MM, *et al*. Maternal exposure to spermicides in relation to certain birth defects. *New Engl J Med* 1987; **317**: 474–8.
 50. Mitchell AA, Rosenberg L, Shapiro S, *et al*. Birth defects related to Bendectin use in pregnancy: I. Oral clefts and cardiac defects. *J Am Med Assoc* 1981; **245**: 2311–4.
 51. Mitchell AA, Schwingl PJ, Rosenberg L, *et al*. Birth defects in relation to Bendectin use in pregnancy: II. Pyloric stenosis. *Am J Obstet Gynecol* 1983; **147**: 737–42.
 52. Andrews EB, Yankaskas BC, Cordero JF, Schoefler K, Hampp S. Acyclovir in pregnancy registry: six years' experience. The Acyclovir in Pregnancy Registry Advisory Committee. *Obstet Gynecol* 1992; **79**: 7–13.
 53. White AD, Andrews EB. The Pregnancy Registry program at Glaxo Wellcome Company. *J Allergy Clin Immunol* 1999; **103**: S362–3.
 54. Herbst AL, Ulfelder H, Poskanzer DC. Association of maternal stilbestrol therapy with tumor appearance in young women. *New Engl J Med* 1971; **284**: 878–81.
 55. Herbst AL, Kurman RJ, Scully RE, Poskanzer DC. Clear-cell adenocarcinoma of the genital tract in young females. *New Engl J Med* 1972; **287**: 1259–64.
 56. Thapa PB, Whitlock JA, Brockman Worrell KG, Gideon P, Mitchel EF Jr, Roberson P, Pais R, Ray WA. Prenatal exposure to metronidazole and risk of childhood cancer: a retrospective cohort study of children younger than 5 years. *Cancer* 1998; **83**: 1461–8.
 57. Kinnier-Wilson LM, Kneale GW, Stewart AM. Childhood cancer and pregnancy drugs. *Lancet* 1981; **2**: 314–5.
 58. Olsen JH, Hertz H, Blinkenberg K, Verder H. Vitamin K regimens and incidence of childhood cancer in Denmark. *Br Med J* 1994; **308**: 895–6.
 59. Parker L, Cole M, Craft AW, Hey EN. Neonatal vitamin K administration and childhood cancer in the north of England: retrospective case-control study. *Br Med J* 1998; **316**: 189–93.
 60. McKinney PA, Juszczak E, Findlay E, Smith K. Case-control study of childhood leukaemia and cancer in Scotland: findings for neonatal intramuscular vitamin K. *Br Med J* 1998; **316**: 173–7.
 61. von Kries R, Gobel U, Hachmeister A, Kaletsch U, Michaelis J. Vitamin K and childhood cancer: a population based case-control study in Lower Saxony, Germany. *Br Med J* 1996; **313**: 199–203.
 62. Mills JL, Fradkin J, Schonberger L, Gunn W, Thomson RA, Piper J, Wysowski D, Brown P. Status report on the US Human Growth Hormone Recipient Follow-Up Study. *Horm Res* 1990; **33**: 116–9.
 63. Fradkin JE, Schonberger LB, Mills JL, Gunn WJ, Piper JM, Wysowski DK, Thomson R, Durako S, Brown P. Creutzfeldt-Jakob Disease in pituitary growth hormone recipients in the United States. *J Am Med Assoc* 1991; **265**: 880–4.
 64. Huillard d'Aignaux J, Alperovitch A, Maccario J. A statistical model to identify the contaminated lots implicated in iatrogenic transmission of

- Creutzfeldt–Jakob disease among French human growth hormone recipients. *Am J Epidemiol* 1998; **147**: 597–604.
65. Billette de Villemeur T, Deslys JP, Pradel A, Soubrie C, Alperovitch A, Tardieu M, Chaussain JL, Hauw JJ, Dormont D, Ruberg M, Agid Y. Creutzfeldt–Jakob disease from contaminated growth hormone extracts in France. *Neurology* 1996; **47**: 690–5.
 66. Buchanan CR, Preece MA, Milner RD. Mortality, neoplasia, and Creutzfeldt–Jakob disease in patients treated with human pituitary growth hormone in the United Kingdom. *Br Med J* 1991; **302**: 824–8.
 67. Tucker MA, Meadows AT, Boice JD, Stovall M, Oberlin O, Stone BJ, Birch J, Voute PA, Hoover RN, Fraumeni JF. Leukemia after therapy with alkylating agents in childhood cancer. *J Natl Cancer Inst* 1987; **78**: 459–64.
 68. Hawkins MM, Kinnier-Wilson LM, Stovall MA, Marsden HB, Potok MHN, Kingston JE, Chessells JM. Epipodophyllotoxins, alkylating agents, and radiation and risk of secondary leukaemia after childhood cancer. *Br Med J* 1992; **304**: 951–8.
 69. deVathaire F, Francois P, Hill C, Schweisguth O, Rodary C, Sarrazin D, Oberlin O, Beurtheret C, Dutreix A, Flamant R. Role of radiotherapy and chemotherapy in the risk of second malignant neoplasms after cancer in childhood. *Br J Cancer* 1989; **59**: 792–6.
 70. Tucker MA, D'Angio GJ, Boice JD, Strong LC, Li FP, Stovall M, Stone BJ, Green DM, Lombardi F, Newton W. Bone sarcomas linked to radiotherapy and chemotherapy in children. *New Engl J Med* 1987; **317**: 588–93.
 71. Glasser DB, Duane K, Lane JM, Healey JH, Caparros-Sison B. The effect of chemotherapy on growth in the skeletally immature individual. *Clin Orthop* 1991; **262**: 93–100.
 72. Roman J, Villaizan CJ, Garcia-Foncillas J, Salvador J, Sierrasesumaga L. Growth and growth hormone secretion in children with cancer treated with chemotherapy. *J Pediatr* 1997; **131**: 105–21.
 73. Mohnike K, Dorffel W, Timme J, Kluba U, Aumann V, Vorwerk P, Mittler U. Final height and puberty in 40 patients after antileukaemic treatment during childhood. *Eur J Pediatr* 1997; **156**: 272–6.
 74. Brennan BM, Rahim A, Adams JA, Eden OB, Shalet SM. Reduced bone mineral density in young adults following cure of acute lymphoblastic leukaemia in childhood. *Br J Cancer* 1999; **79**: 1859–63.
 75. Arikoski P, Komulainen J, Voutilainen R, Riikonen P, Parviainen M, Tapanainen P, Knip M, Kroger H. Reduced bone mineral density in long-term survivors of childhood acute lymphoblastic leukemia. *J Pediatr Hematol Oncol* 1998; **20**: 234–40.
 76. Bruning PF, Pit JM, deJong-Bakker M, van den Ende A, Hart A, van Enk A. Bone mineral density after adjuvant chemotherapy for premenopausal breast cancer. *Br J Cancer* 1990; **61**: 308–10.
 77. Redman JR, Bajorunas DR, Wong G, McDermott K, Gnecco C, Schneider R, Lacher JM, Lane JM. Bone mineralization in women following successful treatment of Hodgkin's disease. *Am J Med* 1988; **85**: 65–72.
 78. Holmes SJ, Whitehouse RW, Clark ST, Crowther DC, Adams JE, Shalet SM. Reduced bone mineral density in men following chemotherapy for Hodgkin's disease. *Br J Cancer* 1994; **70**: 371–5.
 79. Kessler DA, Pape SM, Sundwall DN. The Federal regulation of medical devices. *New Engl J Med* 1987; **317**: 357–66.
 80. The nation: assigning blame; the court vs. scientific certainty. *NY Times* 1999; 27 June: 5.
 81. Institute of Medicine Committee on the Safety of Silicone Breast Implants. Silicone breast implants do not cause chronic disease, but other complications are of concern. *National Academies News* 1999; June 21 (<http://www4.nationalacademies.org/news.hsf/isbn/030906521?OpenDocument>).
 82. The Boston Consulting Group, Inc. (BCG). *Sustaining Innovation in U.S. Pharmaceuticals: Intellectual Property Protection and the Role of Patents*. Boston, MA: BCG, 1996.
 83. Kaitin KI, Manocchia M. The new drug approvals of 1993, 1994, and 1995: trends in drug development. *Am J Ther* 1997; **4**: 46–54.
 84. DiMasi JA. Success rates for new drugs entering clinical testing in the United States. *Clin Pharmacol Ther* 1995; **58**: 1–14.
 85. Chute CG, Panser LA, Girman CJ, Oesterling JE, Guess HA, Jacobsen SJ, Lieber MM. The prevalence of prostatism: a population-based survey of urinary symptoms. *J Urol* 1993; **150**: 85–9.
 86. Sagnier PP, Girman CJ, Garraway M, Kumamoto Y, Lieber MM, Richard F, MacFarlane G, Guess HA, Jacobsen SJ, Tsukamoto T, Boyle P. International comparison of the community prevalence of symptoms of prostatism in four countries. *Eur Urol* 1996; **29**: 15–20.
 87. Chute CG, Stephenson WP, Guess HA, Lieber M. Benign prostatic hyperplasia: a population-based study. *Eur Urol* 1991; **20**: 11–7.
 88. Jacobsen SJ, Girman CJ, Guess HA, Rhodes T, Oesterling JE, Lieber MM. Natural history of prostatism: longitudinal changes in voiding symptoms in community dwelling men. *J Urol* 1996; **155**: 595–600.
 89. Roberts RO, Jacobsen SJ, Rhodes T, Girman CJ, Guess HA, Lieber MM. Natural history of prostatism: impaired health states in men with

- lower urinary tract symptoms. *J Urol* 1997; **157**: 1711–7.
90. Rhodes T, Girman CJ, Jacobsen SJ, Roberts RO, Guess HA, Lieber MM. Longitudinal prostate growth rates during 5 years in randomly selected community men 40 to 79 years old. *J Urol* 1999; **161**: 1174–9.
 91. Jacobsen SJ, Jacobson DJ, Girman CJ, Roberts RO, Rhodes T, Guess HA, Lieber MM. Natural history of prostatism: risk factors for acute urinary retention. *J Urol* 1997; **158**: 481–7.
 92. Stang PE, Osterhaus JT. Impact of migraine in the United States: data from the National Health Interview Survey. *Headache* 1993; **33**: 29–35.
 93. Hu XH, Markson LE, Lipton RB, Stewart WF, Berger ML. Burden of migraine in the United States: disability and economic costs. *Arch Intern Med* 1999; **159**: 813–8.
 94. Maier WC, Arrighi HM, Morray B, Llewellyn C, Redding GJ. The impact of asthma and asthma-like illness in Seattle school children. *J Clin Epidemiol* 1998; **51**: 557–68.
 95. Souetre E, Thwaites RM, Yeardley HL. Economic impact of Alzheimer's disease in the United Kingdom. Cost of care and disease severity for non-institutionalised patients with Alzheimer's disease. *Br J Psychol* 1999; **174**: 51–5.
 96. Turner S, Longworth A, Nunn AJ, Choonara I. Unlicensed and off label drug use in paediatric wards: prospective study. *Br Med J* 1998; **316**: 343–5.
 97. Simons LA, Tett S, Simons J, Lauchlan R, McCallum J, Friedlander Y, Powell I. Multiple medication use in the elderly. Use of prescription and non-prescription drugs in an Australian community setting. *Med J Aust* 1992; **57**: 242–6.
 98. Millar WJ. Multiple medication use among seniors. *Health Rep* 1998; **9**: 11–7.
 99. Sclar DA, Robinson LM, Skaer TL, Galin RS. Trends in the prescribing of antidepressant pharmacotherapy: office-based visits, 1990–1995. *Clin Ther* 1998; **20**: 870–84.
 100. Martinez M, Agusti A, Arnau JM, Vidal X, Laporte JR. Trends of prescribing patterns for the secondary prevention of myocardial infarction over a 13-year period. *Eur J Clin Pharmacol* 1998; **54**: 203–8.
 101. McCaig LF, Hughes JM. Trends in antimicrobial drug prescribing among office-based physicians in the United States. *J Am Med Assoc* 1995; **273**: 214–9.
 102. Schmader KE, Hanlon JT, Fillenbaum GG, Huber M, Pieper C, Horner R. Medication use patterns among demented, cognitively impaired and cognitively intact community-dwelling elderly people. *Age Ageing* 1998; **27**: 493–501.
 103. Hanlon JT, Landerman LR, Wall WE Jr, Horner RD, Fillenbaum GG, Dawson DV, Schmader KE, Cohen HJ, Blazer DG. Is medication use by community-dwelling elderly people influenced by cognitive function? *Age Ageing* 1996; **25**: 190–6.
 104. Harlap S, Davies AM, Haber M, Prywes R, Rossman H, Samueloff N. Patterns of medication in early pregnancy: a preliminary report from the Jerusalem perinatal study. *Adv Exp Med Biol* 1972; **27**: 489–504.
 105. Rosenberg MJ, Waugh MS, Meehan TE. Use and misuse of oral contraceptives: risk indicators for poor pill taking and discontinuation. *Contraception* 1995; **51**: 283–8.
 106. Dahlof C, Boucharard J, Cortelli P, Heywood J, Jansen JP, Pham S, Hirsch J, Adams J, Miller DW. A multinational investigation of the impact of subcutaneous sumatriptan. II: Health-related quality of life. *Pharmacoeconomics* 1997; **11**: 24–34.
 107. Tenover JL, Pagano GA, Morton AS, Liss CL, Byrnes CA. Efficacy and tolerability of finasteride in symptomatic benign prostatic hyperplasia: a primary care study. *Clin Ther* 1997; **19**: 243–58.
 108. Bousquet J, Duchateau J, Pignat JC, Fayol C, Marquis P, Mariz S, Ware JE, Valentin B, Burtin B. Improvement of quality of life by treatment with cetirizine in patients with perennial allergic rhinitis as determined by a French version of the SF-36 questionnaire. *J Allergy Clin Immunol* 1997; **98**: 309–16.
 109. Zhao SZ, Dedhiya SD, Bocanegra TS, Fort JG, Kuss ME, Rush SM. Health-related quality-of-life effects of oxaprozin and nabumetone in patients with osteoarthritis of the knee. *Clin Ther* 1999; **21**: 205–17.
 110. Pariente PD, LePen C, Los F, Bousquet J. Quality-of-life outcomes and the use of antihistamines in a French national population-based sample of patients with perennial rhinitis. *Pharmacoeconomics* 1997; **12**: 585–95.
 111. Zhao S, Yu SS, Dedhiya SD, Zhao WW, Osterhaus JT. Celecoxib improves health-related quality of life in patients with rheumatoid arthritis, Abstract 126, *Pharmacotherapy* 1999; **19**: 497.
 112. Tome MB, Isaac MT. Cost–benefit and cost-effectiveness analysis of the rapid onset of selective serotonin reuptake inhibitors by augmentation. *Int J Psychiatry Med* 1997; **27**: 377–90.
 113. Andersson F, Kartman B, Andersson OK. Cost-effectiveness of felodipine-metoprolol (Logimax) and enalapril in the treatment of hypertension. *Clin Exp Hypertens* 1998; **20**: 833–46.
 114. Neumann PJ, Hermann RC, Kuntz KM, Araki SS, Duff SB, Leon J, Berenbaum PA, Goldman PA, Williams LW, Weinstein MC. Cost-effectiveness of donepezil in the treatment of mild or moderate Alzheimer's disease. *Neurology* 1999; **2**: 1138–45.
 115. Santos R, Cifaldi M, Gregory C, Seitz P. Economic outcomes of a targeted intervention program: the

- costs of treating allergic rhinitis patients. *Am J Man Care* 1999; **5**: S225–34.
116. Testa MA, Simonson DC. Health economic benefits and quality of life during improved glycemic control in patients with type 2 diabetes mellitus: a randomized, controlled, double-blind trial. *J Am Med Assoc* 1998; **280**: 1490–6.
 117. Stampfer MJ, Colditz GA, Willett WC, Manson JE, Rossner B, Speizer FE, Hennekens CH. Postmenopausal estrogen therapy and cardiovascular disease: ten-year follow-up from the Nurses' Health Study. *New Engl J Med* 1991; **325**: 756–62.
 118. Krall EA, Dawson-Hughes B, Hannan MT, Wilson PW, Kiel DP. Postmenopausal estrogen replacement and tooth retention. *Am J Med* 1997; **102**: 536–42.
 119. Zhang Y, McAlindon TE, Hannan MT, Chaisson CE, Klein R, Wilson PW, Felson DT. Estrogen replacement therapy and worsening of radiographic knee osteoarthritis: the Framingham Study. *Arthrit Rheum* 1998; **41**: 1867–73.
 120. Thun MJ, Namboodiri MM, Heath CW. Aspirin use and reduced risk of fatal colon cancer. *New Engl J Med* 1991; **325**: 1593–6.
 121. Rosenberg L, Louik C, Shapiro S. Nonsteroidal antiinflammatory drug use and reduced risk of large bowel carcinoma. *Cancer* 1998; **82**: 2326–33.
 122. Farrow DC, Vaughan TL, Hansten PD, Stanford JL, Risch HA, Gammon MD, Chow WH, Dubrow R, Ahsan H, Mayne ST, Schoenberg JB, West AB, Rotterdam H, Fraumeni JF Jr, Blot WJ. Use of aspirin and other nonsteroidal anti-inflammatory drugs and risk of esophageal and gastric cancer. *Cancer Epidemiol Biomarkers Prev* 1998; **7**: 97–102.
 123. Schreinemachers DM, Everson RB. Aspirin use and lung, colon, and breast cancer incidence in a prospective study. *Epidemiology* 1994; **5**: 138–46.
 124. Stipp D. Scientists use medical-record databases to detect adverse side effects of drugs. *The Wall Street Journal* 1988; 24 March: 33.
 125. Garcia Rodriguez LA, Perez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol* 1998; **45**: 419–25.
 126. Miller E, Blatman B, Einarson TR. A survey of population-based drug databases in Canada. *Can Med Assoc J* 1996; **154**: 1855–64.
 127. van der Linden PD, van der Lei J, Vlug AE, Stricker BH. Skin reactions to antibacterial agents in general practice. *J Clin Epidemiol* 1998; **51**: 703–8.
 128. Evans JM, MacDonald TM. Record-linkage for pharmacovigilance in Scotland. *Br J Clin Pharmacol* 1999; **47**: 105–10.
 129. Glynn RJ, Monane M, Gurwitz JH, Choodnovskiy I, Avorn J. Agreement between drug treatment data and a discharge diagnosis of diabetes mellitus in the elderly. *Am J Epidemiol* 1999; **49**: 541–9.
 130. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease. Comparing administrative data and self-reports. *Med Care* 1997; **35**: 932–47.
 131. Rawson NS, 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.
 132. Young TK, Roos NP, Hammerstrand KM. Estimated burden of diabetes mellitus in Manitoba according to health insurance claims: a pilot study. *Can Med Assoc J* 1991; **144**: 318–24.
 133. Melfi CA, Croghan TW. Use of claims data for research on treatment and outcomes of depression care. *Med Care* 1999; **37**: AS77–80.
 134. Ray WA, Griffin MR, Malcolm E. Cyclic antidepressants and the risk of hip fracture. *Arch Intern Med* 1991; **151**: 754–6.
 135. Garcia Rodriguez LA, Walker AM, Perez Gutthann S. Nonsteroidal antiinflammatory drugs and gastrointestinal hospitalizations in Saskatchewan: a cohort study. *Epidemiology* 1992; **3**: 337–42.
 136. Downey W. Current status of the Saskatchewan Drug Plan. In: Edlavitch S, ed., *Pharmacoepidemiology, Proceedings of the Third International Conference on Pharmacoepidemiology*. Chelsea, MI: Lewis, 1989; 207–9.
 137. A Canadian Database Restored. *SCRIP World Pharm News* 1988; **1302**: 22.
 138. Risch HA, Howe GR. Menopausal hormone usage and breast cancer in Saskatchewan: a record-linkage cohort study. *Am J Epidemiol* 1994; **139**: 670–83.
 139. Risch HA, Howe GR. Menopausal hormone use and colorectal cancer in Saskatchewan: a record linkage cohort study. *Cancer Epidemiol Biomarkers Prev* 1995; **4**: 21–8.
 140. Jick H, Walker AM, Rothman RJ, et al. Vaginal spermicides and congenital disorders. *J Am Med Assoc* 1981; **245**: 1329–32.
 141. Rothman RJ. Spermicide use and Down's syndrome. *Am J Public Health* 1982; **72**: 399–401.
 142. Shapiro S, Slone D, Heinonen OP, et al. Birth defects and vaginal spermicides. *J Am Med Assoc* 1982; **247**: 2381–4.
 143. Mills JL, Harley EE, Reed GF, et al. Are spermicides teratogenic? *J Am Med Assoc* 1982; **248**: 2148–51.
 144. Bracken M, Vita K. Frequency of non-hormonal contraception around conception and association with congenital malformations in offspring. *Am J Epidemiol* 1983; **117**: 281–91.
 145. Linn S, Schoenbaum SC, Monson RR, et al. Lack of association between contraceptive usage and

- congenital malformations in offspring. *Am J Obstet Gynecol* 1983; **147**: 923–8.
146. Harlap S, Shiono PH, Ramcharan S, *et al.* Chromosomal abnormalities in the Kaiser-Permanente Birth Defects Study, with special reference to contraceptive use around the time of conception. *Teratology* 1985; **31**: 381–7.
147. Warburton D, Neugut RH, Lustenberger A, *et al.* Lack of association between spermicide use and trisomy. *New Engl J Med* 1987; **317**: 478–82.
148. Watkins RN. Vaginal spermicides and congenital disorders: the validity of a study. *J Am Med Assoc* 1986; **256**: 3095.
149. Jick H, Walker AM, Rothman RJ. Vaginal spermicides and congenital disorders: the validity of a study. *J Am Med Assoc* 1986; **256**: 3095–6.
150. FDA. Data do not support association between spermicides, birth defects. *FDA Drug Bull* 1986; **16**: 21.
151. Mills JL, Alexander D. Teratogens and “litogens”. *New Engl J Med* 1986; **315**: 1234–6.
152. Holmes LB. Teratogen update: Bendectin. *Teratology* 1983; **27**: 277–81.
153. Neutel CI, Johansen HL. Measuring drug effectiveness by default: the case of Bendectin. *Can J Public Health* 1995; **86**: 66–70.
154. Barbone F, McMahon AD, Davey PG, Morris AD, Reid IC, McDevitt DG, MacDonald TM. Association of road-traffic accidents with benzodiazepine use. *Lancet* 1998; **352**: 1331–6.

8

A View From Regulatory Agencies

DAVID J. GRAHAM

Office of Postmarketing Drug Risk Assessment, Center for Drug Evaluation and Research, Food and Drug Administration, USA

PATRICK C. WALLER

Post-Licensing Division, Medicines Control Agency, London, UK

XAVIER KURZ

Belgian Centre for Pharmacovigilance, Ministry of Public Health, Brussels, Belgium

INTRODUCTION

Drug regulatory agencies exist in most countries, and are accountable for the safety of marketed medicines. They perform a public health function, the ultimate aim of which is to promptly safeguard patients from the adverse effects of medicines. Safety judgments made at the time of authorization need to be revised based on experience—hence the need for proactive monitoring and ongoing risk–benefit analysis.

Spontaneous adverse drug reaction (ADR) reporting is a cornerstone of this work and, although well established, it continues to be developed in response to environmental change (see Chapters 10, 11, and 32). However, the method has important limitations and formal pharmacoepidemiology studies are of great value in regulatory decision making. Their role has

steadily increased over the years as regulators have recognized the importance of using comparative data and powerful methodologies. Modern databases have facilitated the availability of such data and provided resources for valid and timely investigations of drug safety issues.

Postmarketing regulatory decision making involves bringing together various types of data and often requires difficult judgements based on conflicting or inadequate data. The key output is communication with health professionals and patients of the necessary measures to promote safer use of medicines. International co-operation has increased greatly during the last decade and has important benefits. Technological advances are providing an opportunity to exchange information, making processes more efficient and facilitating common practices.

The views expressed are those of the authors and not necessarily those of their respective employers.

THE NEED FOR EFFECTIVE DRUG SAFETY PROGRAMS

The premarketing drug development process has been divided into three phases involving exposure to humans.¹ Phase 1 generally involves small numbers of healthy volunteers who receive one or a few doses of a drug product for purposes of determining dose ranges and patterns of drug metabolism. During phase 2, small numbers of patients are treated, with the goal of defining the typical daily dose as well as of determining the pharmacokinetics and the most common side-effects of the drug under study. In phase 3, larger numbers of patients are treated in the setting of randomized clinical trials that are designed primarily to assess product efficacy for a specified indication, and to further characterize the safety profile of the drug.

Phase 3 trials often range in size from a few hundred to a few thousand patients, who may be treated for durations varying from a single dose to several months. The median number of patients enrolled in premarketing studies of new active substances was estimated in the late 1980s to be around 1500.² For modern biotechnology products intended for use by specialists for the treatment of targeted smaller populations, this number is almost certainly lower. Patients enrolled in phase 3 trials are usually screened to exclude those not meeting certain narrowly defined criteria. The purpose of such screening is to reduce the presence of potential confounding factors that might cloud the analysis of efficacy. For example, patients with underlying renal or hepatic dysfunction or those taking other medications are often excluded. Additionally, children and older aged subjects are usually not highly represented. As a result of these design elements, phase 3 trials have limited ability to detect serious, but uncommon adverse reactions.³⁻⁷

Because of sample size limitations, most phase 3 trials lack the statistical power to observe adverse reactions occurring at rates of 1 per 1000 treatment courses or less (see also Chapter 3). Also, because of the typical shorter duration of these trials, adverse reactions dependent on chronic or longer term exposure or those with long latency to onset cannot usually be observed. The tendency for

patients with complicated medical histories or those taking multiple other medications to be excluded from phase 3 studies effectively precludes the possibility of detecting clinically relevant interactions between the study drug and a variety of disease states or other medications. Finally, because of the relatively narrow and specific selection criteria of trial participants, adverse reactions arising in patient subgroups or under less common circumstances cannot be observed.

International experience relating to the withdrawal of medicines for safety reasons illustrates the inherent limitations of phase 3 trials to identify serious, uncommon adverse drug reactions (ADRs).^{8,9} Studies indicate that, over a period of about 20 years, market withdrawals occur at a frequency of about 3–4%. In general, only after exposure to larger numbers of patients in a broader array of clinical settings than those studied premarketing did the reactions that led to a major safety concern become clinically apparent. In some cases (e.g., encainide, flequinan), excess mortality was discovered in the setting of large clinical trials during the postmarketing period.

Many withdrawals for safety reasons have occurred simultaneously on a worldwide basis, but there have been some notable exceptions. For example, terfenadine was withdrawn in the USA in 1998 because of its effect in prolonging the QT interval, but it remains available with restrictions in the European Union (EU). In the past two years, hepatotoxicity associated with troglitazone (an antidiabetic), trovafloxacin (a quinolone antibiotic), and tolcapone (an anti-Parkinsonian) has led to withdrawal of these drugs in the EU but they have remained available with labeling safeguards in the USA.* The reasons for these differing decisions have not been systematically explored but could include differences in perceptions of risk and of the effectiveness of potential safeguards. Other elements to consider are variations in clinical practice, in the nature of the regulatory powers available to the relevant authorities, and in the availability of alternative products for the same indication.

* Troglitazone was removed from the US market after this chapter went to press.

Although withdrawals for safety reasons are fairly uncommon, new adverse reactions are discovered postmarketing for most drugs, resulting in a need to update the product information. Thus, our understanding of a drug's safety profile evolves and changes as we learn more about it during the postmarketing phase of its development cycle. The drug development process does not end the day the product is approved for marketing, but continues throughout the postmarketing period. However, the first few years of marketing are a particularly important time and in some countries this is recognized by a formal intensive monitoring scheme for new drugs (e.g., the use of the black triangle symbol in the UK).

During this period, it is vital that new information regarding drug product safety be collected, analyzed, and acted upon. First and foremost, this information is important to the public. People naturally desire to be as fully informed as possible of potential risks associated with the drug products they consume. An important function of a drug regulatory authority in this regard is to ensure that important new safety information is identified as early as possible and then conveyed in an accurate and understandable manner to the public and to health practitioners. This information is important to the process of updating a product's labeling, to refining the conditions for safe prescribing, and for the combined activities of risk management and regulatory decision making.

In a regulatory environment, the goal of postmarketing surveillance is the evaluation and prevention of drug hazards based on the collection, analysis, and communication of drug risk information. Pharmacoepidemiology has an essential role in this setting to help identify and evaluate adverse effects early in the postmarketing phase. Especially for those effects of greatest severity or with the greatest population impact, additional goals become important, namely, estimation of incidence, risk quantification in comparison to therapeutic alternatives, assessment of a causal relationship, elucidation of the determinants contributing to their occurrence (including patterns of drug use), and conveyance of relevant risk estimates to drug regulators.¹⁰ It is with these latter activities that the tools of pharmacoepidemiology

uniquely contribute to postmarketing drug safety efforts.

Epidemiologic data must be considered in the context of other sources of data relating to the occurrence of serious adverse reactions. Frequently, clues to potential safety concerns arise from the clinical pharmacology of a drug product (see also Chapter 4). A drug's pharmacokinetic or pharmacodynamic properties may provide advance notice of areas to focus attention upon in the postmarketing phase. For each aspect of a product's pharmacokinetic journey through the human body, from absorption through distribution, metabolism, and elimination, opportunities exist for ADRs to arise.¹¹ An especially important source of ADRs to emerge during the 1990s has been that of drug-drug interactions, primarily involving the cytochrome P450 system of isoenzymes.^{12,13}

Safety concerns may also be signaled by observations during phases 2 and 3 of premarketing development. A drug that induces frequent or severe liver transaminase elevations in clinical trials (even though they return to normal with discontinuation of therapy) may be a candidate to monitor for the occurrence of hepatitis and acute liver failure if the product is marketed. Depending upon the nature of the safety concern identified premarketing, as well as the anticipated indication and extent of use postmarketing, a phase 4 commitment to conduct additional focused safety studies during the postmarketing period might be agreed to by a product's manufacturer.

Once a drug is marketed, evidence of potential safety issues may be obtained from sources such as the medical literature (case reports, clinical trials, epidemiologic studies) or spontaneous case reports submitted to regulatory authorities (see Chapters 10 and 11). New safety information may be developed by the manufacturer or come from regulatory authorities and postmarketing experience in other countries.

When data from these multiple sources are properly integrated, analyzed, and interpreted, they become information which can help guide drug regulatory policy. This process of discovery requires a multifaceted interdisciplinary approach that draws on the fields of medicine, clinical pharmacology, and toxicology, and relies upon the methodologies

and tools of epidemiology and biostatistics. While it is frequently the case that the focus of this process is on new molecular entities that have been marketed for less than three years, older drugs also must be screened for important and previously unrecognized adverse effects. Ultimately, the timing of when a signal of a potentially new and serious adverse reaction is identified will depend upon the extent of product use in the population, the incidence rate of the adverse reaction, and the rate of reporting of that reaction to the literature, manufacturers, and regulatory authorities.¹⁴

ELEMENTS OF A DRUG SAFETY PROGRAM

There are a number of components that contribute in varying degrees to national postmarketing drug safety programs.

DRUG USAGE DATA

Drug usage data is a valuable source of information important to drug safety surveillance and pharmacoepidemiology. It is of particular importance in the analysis and interpretation of spontaneous ADR data (see below).

The availability of drug usage data and the access to them vary greatly from country to country. In many countries, sales figures are the main source of information on drug usage, whether they are available from compulsory notification to authorities (e.g., within the framework of price–volume contracts) or from specific requests to companies. Sales figures are useful to estimate the total number of days of treatment in a defined period but, except in special conditions, they cannot provide accurate information on the number of treatments or the number of patients treated.

In some countries, the concept of a defined daily dose (DDD) is used to improve the utility of sales data. The DDD is a technical, arbitrary, unit of measurement defined as the average daily dose for an adult of 70 kg using the drug in its main indication (see Chapter 29). Using DDDs allows the summation of the quantities consumed for a

drug existing in various physical units (e.g. grams, milliliters), preparations, or package sizes. It also allows the summation of consumption data for a drug used differently in various settings (e.g., hospital wards, regions, and countries) or for several drugs belonging to the same therapeutic class. Another advantage is the possibility of comparing the usage of a drug or class of drugs across these various settings, and to assess time trends in consumption. The DDD methodology is therefore used for an aggregate estimation of the consumption of a drug, to provide comparable denominators of exposure, and to evaluate the impact of regulatory decisions on drug usage. For these purposes, it does not require assumptions about patterns of treatment, which may vary from setting to setting. However, the number of patients treated cannot be estimated with accuracy.

Prescription data are also useful for estimating the number of treatment courses of marketed prescription medicines and for providing information on treatment modalities (see also Chapter 24). They become increasingly available to regulatory authorities with computerization of pharmacies and the collection of prescription records into national databases. National and international drug usage data are also available from commercial data providers (e.g., IMS Health).

In the US, two database resources are used routinely—the National Prescription Audit Plus (NPATM) and the National Disease and Therapeutic Index (NDTITM). The NPATM provides national estimates of prescription drug use in the US, based on data collected from over 20 000 computerized chain, independent, grocery, and mail-order pharmacies nationwide. These data can be used to track prescribing trends over calendar time and for calculation of ADR reporting rates. The NDTITM database is based on a prescription audit, from a sample of US physicians. It provides information on the age and gender distribution of patients using mainly outpatient drug products, as well as information on the duration of treatment course, indication for use, and prescriber specialty. Such data are frequently used by US regulatory pharmacoepidemiologists to help understand the population context within which ADRs occur.¹⁵

When NDTI™ data are used with data from NPA™, estimates of the person time of exposure in the population can be derived. One can also estimate exposure in particular age and/or gender strata. However, because both NPA™ and NDTI™ collect data on prescription events rather than longitudinal data on patients, it is often difficult or impossible to estimate the actual size of the population receiving a drug product based on these data alone.

The situation in the UK is similar, with the most important source of usage data being derived from the Prescription Pricing Authority. Since these data are used for reimbursement purposes with the National Health Service they are generally complete but, as with the US data sources described above, they do not provide an accurate denominator of the numbers of patients exposed.

None of these sources provides estimates of inpatient hospital drug use. The Provider Perspective, another IMS Health database, produces estimates of bulk drug purchases by hospitals and other institutional health care providers. However, these data are difficult to translate into estimates of the number of patients treated with parenteral products as hospital inpatients. Other databases are being explored which might be helpful in the assessment of drug safety issues related to products used primarily in the hospital inpatient setting.

ADVERSE REACTION CASE REPORTING

One of the cornerstones of a comprehensive safety monitoring program is spontaneous ADR reporting, which relies on the systematic review of reported cases of suspected ADRs (see also Chapters 10 and 11). In many developed countries, suspected ADR cases are reported to the national drug regulatory authority, where they are reviewed and entered into a computerized database to facilitate future aggregate analyses and report retrieval.^{16–18}

Relevant standards have been agreed upon under the process of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).¹⁹ These include use of a common ADR

coding terminology and common data elements for individual case reports. This will help facilitate electronic submission of ADR data by pharmaceutical companies and the electronic exchange of ADR case reports between national regulatory authorities throughout the world.

Cases of suspected ADRs are reported by health professionals either directly to the regulatory authority or to the manufacturer, which is legally obliged to transmit them to regulatory authorities within defined timeframes (normally 15 days for a report of an ADR considered to be serious).^{4,20} In some countries, notably the USA, patients may also report ADRs to the regulatory authority. In some other countries, they may report to National Poison Centers that liaise as necessary with the regulatory authority.

Substantial emphasis is placed on systematic review of spontaneous ADR reports for the purposes of identifying unrecognized potential hazards and gaining new information about established ADRs. There is a particular focus on ADRs with serious outcomes and ADRs not currently mentioned in the product information. A “serious” outcome is defined as one resulting in death, or which is life-threatening or leads to hospitalization, disability, or a congenital abnormality.²¹

Inherent in any spontaneous adverse reaction reporting system is a monitoring process that recognizes the dynamic nature of data. A prerequisite for an effective system is a database in which new reports can be entered speedily, prioritized according to their importance, and from which data can be retrieved and analyzed in a variety of ways. Regular and systematic review of the database is necessary, and is most commonly achieved by reviewing individual drugs or products looking for reactions of potential concern.

There are a number of positive attributes to national passive case reporting systems. They cover all marketed prescription drugs from the day they first appear in the marketplace. This permits the signaling of potential new reactions early in the postmarketing phase. For example, within four months of the market introduction of temafloxacin in the USA, the FDA received the

first reports of hemolytic anemia and renal failure associated with use of this drug. Two months later, with continued accrual of new cases, temafloxacin was withdrawn from the market worldwide.²² This system was designed to quickly signal rare, serious ADRs once a product reaches the market. Temafloxacin is one of numerous examples of how such systems have performed this task well.

On a per report basis, and considering the breadth of drug coverage, passive case surveillance systems are relatively inexpensive. These systems essentially leverage the eyes and ears of health care professionals and consumers throughout the world in the effort of case finding. This type of surveillance system is especially well suited to the detection of extremely rare or unusual ADRs.¹⁶ Reactions such as acute liver failure, severe skin disorders, aplastic anemia, and hemolytic anemia are generally rare in the population and their occurrence often prompts some exploration for potential exposure to a drug or toxin. In settings where sufficient numbers of cases with usable data are reported, risk factors and at-risk subgroups can also be identified.

There are a number of important limitations to passive case reporting systems. Case reports may lack important clinical details or supporting materials such as laboratory results. Efforts to obtain this information from the primary reporter may be unsuccessful. A more serious limitation is that most adverse reactions are not reported, a phenomenon referred to as underreporting. The process of reporting an ADR may be thought of simplistically as a series of three barriers, each of which must be overcome for a case report to be available for analysis.

The first barrier is that of diagnosis. To be of value, an ADR must be diagnosed, that is, a label applied to it that classifies or specifies what the reaction is or appears to be. The diagnosis may range in specificity from a symptom or clinical sign to a precise, pathologically confirmed disease entity. An ADR can affect any or multiple organ systems in a variety of ways, resulting in a wide variety of potential presentations,¹⁴ contributing further to difficulty in diagnosis. The low specificity of some diagnoses can effectively amount to misclassification when it comes to the analysis of

case reports. If a case of drug induced hemolysis with secondary jaundice is reported as “jaundice” or “hyperbilirubinemia”, the reaction “hemolysis” may be missed. This report might not be included in an analysis focusing on hemolysis because the diagnosis of hemolysis was not made or else was not included in the report.

The second barrier to ADR reporting is that of attribution. A disease process may be diagnosed correctly, but if it is not attributed to a drug exposure, the adverse reaction probably will not be reported. A major impediment to physician or patient attribution of disease signs and symptoms to a drug exposure is the fact that ADRs frequently mimic or resemble naturally occurring disease processes. Additionally, newly developing complaints may be incorrectly attributed to ongoing underlying disease processes rather than to a particular drug exposure.²³ Thus, if an ADR is identical to a disorder occurring at an appreciable rate in the population (e.g., myocardial infarction), or if the ADR is identical to the underlying disease for which the drug is prescribed (e.g., asthma), the diagnosis of an ADR will often not be made. The situation is still more complex when occurrence of a complication or a consequence of the underlying disease can be misdiagnosed as an ADR to a newly introduced drug, and vice versa. The difficulty in dissociating these two factors may lead to failure to recognize severe adverse effects, or, conversely, to misinterpretation of disease complications as adverse drug effects. Examples of associations that illustrate these difficulties are bone pain and diphosphonates,²⁴ reperfusion arrhythmia and thrombolytics,²⁵ hepatic failure and fialuridine,²⁶ and Churg–Strauss syndrome and leucotriene receptor antagonists.²⁷

The final barrier to reporting is the physical act of filing a report of the suspected adverse reaction with a drug sponsor or a regulatory authority (registration of the event). There are many potential reasons why a health professional might not report an ADR. Some years ago Inman described the “seven deadly sins” leading to the failure by physicians to report an adverse effect (complacency, fear of litigation, guilt, ambition to publish, ignorance, diffidence, and lethargy).²⁸ Many studies have been published that address

this final barrier to reporting, almost invariably showing high levels of underreporting. For example, two US surveys found that the reporting rate for serious recognized ADRs in Rhode Island was below 3%,²⁹ and that physicians in Maryland filed reports of serious reactions in only 8–13% of instances.³⁰

In situations where diagnosis or attribution occurs at low rates, the magnitude of underreporting can be much greater. In the UK, an increase in asthma deaths among patients using pressurized inhalers was largely unrecognized and unreported, as was the occurrence of the oculocutaneous syndrome with praxolol.^{28,31} The reporting efficiency in each of these instances was substantially below 1%. Low ADR reporting efficiency has important implications for how one uses and interprets spontaneous case reports data in a population or public health context.

Recent research across Europe has focused on physician attitudes to reporting and found that uncertainty about attribution and a perception that certain ADRs are trivial or well recognized are the most common reasons for nonreporting.^{32–34} Lack of awareness about ADR reporting is another underlying factor, but competing pressures on reporters' time is probably even more important and needs to be counteracted by specific measures to stimulate reporting. Some authors have tried to identify incentives to reporting rather than barriers.³⁵ Contribution to scientific knowledge, unexpectedness, severity of the reaction, and novelty of the suspected drug were the most frequently retrieved reasons for reporting.

Regulatory authorities have adopted various approaches to stimulate reporting. Broadly, these fall into four categories:

1. educational efforts directed at professional groups (e.g., medical students and junior doctors);
2. broadening the base of reporters by targeting specialists in particular therapeutic areas (e.g., HIV medicine, oncology) for which reporting tends to be low;
3. facilitating the process of reporting by making reporting forms easy to complete and widely

available (including electronically via the internet); and

4. improved feedback to reporters (e.g., through bulletins).

POPULATION-BASED DATA RESOURCES

Because of the limitations of spontaneous ADR reporting, particularly in respect to measuring frequency and establishing causality, there is a critical need for population based resources. While these could be used for hypothesis generation, their principal value lies in rapid hypothesis strengthening and formal hypothesis testing.³⁶ Increasingly, therefore, regulatory authorities are using such resources in the investigation of important safety issues. In the USA, the FDA maintains a Cooperative Agreement Program in Pharmacoepidemiology that provides access to claims based and record linked data from a variety of population based health care settings.³⁷ The principal databases currently represented in this program include the General Practitioners Research Database (GPRD) from the UK (see Chapter 23), the Harvard Pilgrim Health Plan (a health maintenance organization) (see Chapter 17), Tennessee Medicaid (see Chapter 19), and United HealthGroup (multistate managed care) (see Chapter 18).

In the UK, the GPRD is regarded as a vital resource for pharmacoepidemiology (see Chapter 23). This database has been shown to provide information of sufficient quality for research purposes³⁸ and in the past decade has been principally used for pharmacoepidemiology. The GPRD is now the largest source of longitudinal patient data in the world. During 1999, its management was taken over by the UK Medicines Control Agency, and a major investment in the infrastructure is being made in order to ensure that its potential is maximized.

The key elements of such databases are that they provide longitudinal followup of large numbers of patients with computerized information on outpatient prescriptions linked to diagnoses and hospitalizations. Also important is the capability to access primary medical records for outcome validation and abstraction of other information

relevant to a particular safety issue. Such databases can be used to more accurately describe the demographics of drug usage and can demonstrate patterns of coprescribing with other medications, which is especially important to the study of drug–drug interactions.³⁹ The data can also be used to determine the duration and pattern of product usage, thereby permitting one to create a life table describing the proportion of patients who remain on treatment for varying lengths of time. This type of information can be used to model prescription data and therefore to provide estimates of the number of patients using a particular drug or product.

Depending on the safety issue and related specifics of clinical pharmacology and medical setting, these databases can be used to assemble case series, and for the performance of retrospective cohort and/or case–control studies where the objective may be to estimate incidence rates or relative risks, or to identify important risk factors.^{40–42} An important advantage of such databases is that information on drug exposures and health events/outcomes is collected prospectively and in an unbiased fashion over time. Furthermore, it is often possible to study emerging safety issues rapidly and efficiently.

RESPONSES TO SIGNALS

CASE REPORTS AND CASE SERIES

The most common use of spontaneous case reports is for the updating of product information to include newly identified ADRs. In most countries performing pharmacovigilance activities, all serious unlabeled reports from manufacturers and all reports submitted directly by health professionals and consumers are reviewed by experienced safety evaluators.¹⁶ If a report is incomplete or important additional information is needed, followup with the reporter is undertaken. A routine labeling change may be based on a single well documented and compelling report, although typically there will be several well documented and unconfounded reports, with a variable number of reports of lower quality as supporting evidence. A formal causality

assessment algorithm is usually not followed in this process due to the lack of flexible, simple, and reliable algorithms. However, reports are usually classified qualitatively as stronger or less strong in their support of a particular signal.

Occasionally, large numbers of case reports for a given drug–ADR combination will be reviewed for purposes of identifying potential risk factors or to understand better the clinical expression of that ADR. A review of 121 reports of seizure with alprazolam use found that 58% occurred within 24–48 hours of the last dose, suggesting the possibility of drug withdrawal.⁴³ In another example, review of 95 cases of hemolytic anemia with temafloxacin showed that the median time to symptom onset was 6–7 days, compatible with an immune-mediated mechanism.²² Among cases following the first dose, seven of ten had prior exposure to fluoroquinolone antibiotics, compared to 11 of 85 among cases occurring beyond the first dose, further supporting the possibility of an immune mediated mechanism.

A further consideration is the complexity which is due to the large number of concomitant drugs taken by many patients presenting with an ADR, especially by the elderly. This factor can considerably obscure the overall picture. In community studies, drug interactions—broadly defined as the change of the effects of one drug by the presence of another drug—were found to occur at a rate of 2–4% of treated patients.⁴⁴ The suspicion of a potential drug interaction in a case report leads to several difficulties. In many instances, the data provided on drugs taken concomitantly (e.g., dosage and dates of administration) are insufficient to evaluate the possibility of an interaction. A large amount of time is also usually required to locate and interpret drug interaction information, although this time may be reduced by the use of computerized software programs. Moreover, evidence on the clinical outcome of pharmacokinetic interactions is often lacking and much of the available information is based on case reports or small case series. As there is a large variability in patients' response to the presence of potentially interacting drugs, a further limitation may be a lack of generalizability of the information to a wider population of users.

AGGREGATE ANALYSIS OF CASE REPORTS

One can also analyze case report data in an aggregate fashion, for purposes of signal identification or refinement. The proportional distribution of an ADR or group of ADRs within relevant strata of exposure (e.g., gender, age group, dose level, duration of use) for one drug can be compared with others in the same pharmacologic class, or with drugs from different classes used for the same indication. Such proportional distributions, while not informative regarding event incidence, can be useful in highlighting potential problem areas for a particular drug. A review of ADRs reported with amoxicillin-clavulanic acid found that 80% of liver related ADRs were reported in patients age 40 and older. At the same time, only 24% of product use was estimated to be in this older age group, suggesting that age may be a risk factor for hepatotoxicity with this drug product. At least one epidemiologic study has found an association between hepatic ADRs and older age among amoxicillin-clavulanic acid users.⁴⁵

The use of reporting rates is another aggregate approach that is sometimes helpful in the process of signal development or refinement, particularly when comparing drugs within the same pharmacologic class or drugs prescribed for the same indication. With this approach, the number of cases reported with a given ADR serves as the numerator, with a denominator derived from an estimate of product usage, usually the number of outpatient prescriptions. However, reporting rates can be difficult to interpret. Because of underreporting, they cannot be viewed directly as incidence rates. Other factors may further complicate the interpretation of reporting rates. These include differences in the market age of drug products being compared and publicity effects relating to a given drug-ADR association, to name a few.

It is also possible to perform aggregate analyses using the data from case reports that are stored electronically in the ADR database. The data from case reports describing specific drug-ADR combinations can be downloaded into individual datasets for further analysis. These data can be

combined with prescription data to calculate reporting rates. It is possible to compare one drug with another used for the same indication, or to compare one drug class with another. The age and gender distribution of ADR reports can also be compared with that found in drug usage or population based healthcare databases, and subgroup reporting rates can be explored. Evidence of important differences between two drugs or drug classes, or within particular subgroups, may lead to further investigation in a population based setting. Of course, given the known limitations of these data, such analyses must be conducted and interpreted with great caution.

ANALYTIC USES OF CASE REPORTS

In situations where the ADR is a well defined outcome for which there are population based estimates of background incidence available, spontaneous case reports can be used to perform an “observed to expected” analysis.⁴⁶ This method is best applied in settings where the outcome of interest is rare, such as acute liver failure or aplastic anemia, rather than more common outcomes, such as myocardial infarction. The “expected” number of events is obtained as the product of the population based incidence rate for the event multiplied by the person time estimate of the population exposure to the drug under consideration. This number is compared to the number of “observed” cases, that is, the number reported. Because of underreporting, the *a priori* expectation is that the “expected” number will far exceed the number that has been reported. However, in situations where the number reported is close to or exceeds the “expected” number, one has a strong signal of a problem. This approach was used in the study of Guillain-Barré syndrome following receipt of serum-derived hepatitis B vaccine,⁴⁷ in the evaluation of felbamate and aplastic anemia,⁴⁸ and, more recently, in the FDA’s assessment of bromfenac and acute liver failure.⁴⁹

Under certain conditions, spontaneous reports can be combined with population based data to perform an analysis similar to that of a nested case-control study. Cases and noncases are all

exposed to the same drug of interest. The “cases” are those reported spontaneously as ADRs while the “controls” are obtained from a population based source such as a health insurance plan or an HMO. The dichotomized exposure measure being tested may be dose level, duration of use, or some patient-specific covariate. The proportional distributions of cases and noncases are calculated with respect to the categories of exposure, and, from this, a proportional odds ratio is derived. This approach was used to demonstrate that doses of alprazolam above 4 mg d⁻¹ appeared to confer a substantial increase in risk of withdrawal seizure compared to lower daily dosages in the setting of abrupt discontinuation of the drug.⁴³ Recently, Moore *et al.* used voluntary ADR cases reported to the French national database to perform a “case–noncase” analysis.⁵⁰ In this setting, “cases” and “noncases” were obtained from the entire ADR database. Odds ratios were calculated, comparing exposure to particular drugs of interest to all other drugs combined in the database. A similar approach, developed by the UK Medicines Control Agency (Proportional Reporting Ratios), is described elsewhere in this text (see Chapter 11).

OTHER SIGNALING STRATEGIES

Recent advances in information technology have opened new possibilities for the detection of previously unidentified ADR signals based on spontaneous case reports. An approach under development at FDA and WHO applies Bayesian techniques to the matrix of all drugs and all ADR coding terms in the spontaneous reports database to highlight drug–ADR combinations occurring in the database with greater than expected frequency^{51,52} (see also Chapter 32). Experience with this technique has shown promise thus far. In the course of validation testing, the method has successfully identified many known signals that were previously identified, evaluated, and verified. A potential advantage of this technique over current approaches is that it is completely automated and can screen all drug products simultaneously. With additional refinement, it is possible that this tool will be useful in pharmacovigilance,

helping to prioritize signals and guide safety reviewers to focus on those issues that are most important to the public health.

Regulatory authorities adopt a proactive and systematic approach to signal generation. Although many signals are identified, relatively few turn out to require major investigation. Judgments have to be made, often based on limited evidence, as to whether or not there is an issue in need of attention. There are four key factors that help determine whether or not a particular signal is worth further investigation:

1. the *strength* of the signal,
2. whether or not the issue or some aspect of it is *new*,
3. the clinical *importance* as judged by the seriousness of the reaction and severity of the cases, and
4. the potential for *preventive* measures by the regulatory authority (signified by the acronym SNIP).

The SNIP criteria are useful in identifying which signals to pursue in greater depth, by bringing together all the relevant evidence available, and, sometimes, by designing specific studies to further investigate it. Consideration is given at an early stage to the possible outcomes and specifically as to how risks might be minimized.

USE OF POPULATION BASED DATA RESOURCES

Spontaneous case reports have great utility in signaling the existence of potential drug safety issues and frequently provide sufficient clinical information upon which to base labeling changes and other interventions such as informational letters to healthcare professionals. However, these reports do not permit an accurate estimation of incidence or relative risk. In some situations, it is important to understand the population impact and risk associated with a particular drug exposure. This requires formal epidemiologic approaches using data obtained from population based sources (see Part III of this book). In these

settings, the ADR of concern typically carries substantial morbidity or else is life threatening.

The Cooperative Agreement Program mentioned above is one means available to FDA for conducting epidemiologic postmarketing safety studies in the general population. Other approaches involve collaboration with interested professional organizations or else requesting that the company marketing the drug product of interest perform such studies.

Metformin, a biguanide oral hypoglycemic agent, was marketed in Europe and Canada during the 1970s and was approved in the US in 1995. The later approval in the US was related primarily to questions regarding the potential risk of lactic acidosis. This concern arose because phenformin, another biguanide, had been withdrawn from the US market in 1977 because of its association with lactic acidosis. Given the seriousness of the ADR and the anticipated use of metformin, the FDA used its Cooperative Agreement Program to obtain a population based estimate of the incidence rate of lactic acidosis in metformin users.⁵³

A retrospective cohort study was performed using administrative databases from the province of Saskatchewan, Canada.⁵³ All subjects filling at least one prescription for metformin between 1980 and 1995 were enrolled in the cohort contributing 22 296 person years of exposure. Review of computerized hospitalization claims for "acidosis" and subsequent medical record review identified two cases of lactic acidosis within the cohort, for an incidence rate of nine per 100 000 person years of metformin use. Of note, both cases had other medical conditions, each independently capable of causing lactic acidosis. These data provided additional information suggesting that the rate of lactic acidosis was relatively rare and substantially below that previously estimated with phenformin.

In another example, studies were conducted in both the USA⁵⁴ and UK⁵⁵ to investigate the spontaneously reported association between the dose of lipase containing pancreatic enzyme supplements and fibrosing colonopathy. Both were case-control studies with cases of histologically confirmed colonopathy being identified by survey of all accredited cystic fibrosis care centers in the US and from a national register in the UK.

Controls were randomly selected from among cystic fibrosis patients treated with enzyme supplements but without colonopathy. Medical records and/or interviews were reviewed to obtain information on dose and type of supplement used as well as other potential risk factors. Both studies found significant increases in risk of fibrosing colonopathy with increasing daily dose of lipase, and provided support for recommendations that the maximum daily dose of enzyme supplementation should be 10 000 units of lipase kg⁻¹.

USE OF POSTMARKETING SAFETY DATA IN REGULATORY DECISION MAKING

Regulatory decisions relating to postmarketing safety are made within a complex matrix of considerations. This matrix includes the medical importance and utility of the drug in question, the drug's extent of usage, the severity of the disease being treated, the drug's efficacy in treating this disease, the availability of other products to treat the same disorder, the product's mechanism of action, and clinical pharmacologic features of the drug such as mode and site of metabolism. This multidimensional matrix also includes the severity of the ADR, its frequency and incidence in the relevant population, risk factors contributing to the ADR's occurrence, and the clinical context in which the ADR arises.

The focus of regulatory decision making is to reduce or eliminate health risks so as to improve the balance of benefits and risks.¹⁶ There are no fixed formulae for the regulation of drug safety risks; each case is unique. In this context, the role of pharmacoepidemiology is to provide scientific support to help guide regulatory decision making and make it as evidence based as possible. Regulatory authorities respond to important hazards by gathering all the relevant evidence and undertaking a critical risk-benefit review. In the EU there is another dimension, which is the need for consensus across international boundaries. This is achieved through a scientific committee, known as the Committee for Proprietary Medicinal Products (CPMP), which includes representation from

all Member States. The CPMP has several specialist groups including a Pharmacovigilance Working Party, which is the EU forum for detailed debate of postmarketing safety issues.

As indicated above, drug withdrawal from the market because of safety concerns is relatively unusual and, for most issues, the regulatory decision is to amend the product information or to take no action other than continued monitoring. Various sections of the product information may be amended, particularly those specifying indications, contraindications, warnings/precautions, interactions, and adverse reactions. For serious hazards, a “Dear Healthcare Professional” letter might be required to ensure the dissemination of important new safety and/or prescribing information. In special circumstances, a restricted distribution program might be instituted or formal postapproval safety studies requested.

Provision of information to drug users is essential to support the safe and effective use of medicines. The key requirements for a successful drug safety communication are that it should be specific for the target audience, easily understandable, open and informative, and well balanced, placing the issue in an appropriate context.

It is particularly important in any communication about drug safety that essential information is clearly conveyed and not obscured by other less important information. The key facts and recommendations have to be worded unambiguously and placed in a prominently early position, if necessary with use of highlighting.

In the EU, patient information leaflets are now mandatory for all medicines. The key principles with patient information are that it should, in substance, be the same as the information provided to health professionals but presented in language that a lay audience can understand. The aim of patient information is to reinforce the main issues that should be discussed between health professionals and patients, and it should not make statements that could interfere with that relationship. The focus on written information for patients is likely to increase; maximizing its potential value will be an important challenge for regulatory authorities around the world.

After a decision has been made and a course of action taken, continued monitoring and followup are important as a form of feedback to the decision makers. If followup evidence suggests that the safety problem or conditions contributing to it persists despite a particular regulatory intervention, additional regulatory action might be indicated to successfully manage the drug risk.

FUTURE DIRECTIONS

For many drug safety issues, public health responsibilities require that a regulatory decision be made in the face of imperfect and incomplete information. An important area for future development in pharmacoepidemiology relates to finding ways to meet these information needs in an efficient and timely manner, so as to reduce the level of uncertainty regarding the risk component of a drug risk–benefit evaluation.

Spontaneous case reports are an essential aspect of most drug risk assessments, and the case reports of greatest importance to decision makers are those describing serious adverse health effects. Underreporting creates uncertainty because it masks the magnitude of an ADR in the population and makes it difficult to measure the effect of regulatory interventions. Incomplete reporting contributes to uncertainty because important relevant information will often be absent from the report. Efforts to enhance both the quantity and quality of reported cases of suspected serious ADRs would help to reduce uncertainty, and potentially improve the effectiveness of drug risk management.

Another area where informational uncertainty is great relates to the use of drug products in hospitalized patients. There is a need for nationally representative patient-specific drug use data from the hospital inpatient setting. Were such data linked to hospital procedures and patient outcomes, it would represent a valuable addition to the information armamentarium. In a related manner, there is also a strong need for support of the development of surveillance systems for ADRs in hospitals based on recent experiences.^{56–58}

A continuing challenge will be finding ways to make use of advancing technology to process, access, and analyze large numbers of drug safety data. The use of data mining techniques for signal detection within the spontaneous case reports database is a current example.^{51,52} Refinements in this methodology will allow screening of subsets of the database stratified by age, gender, or seriousness of the ADR, and may also be useful in the identification of previously unrecognized drug–drug interactions. It is possible that such methods could also be applied to screening within claims based longitudinal databases.

Active surveillance is another means by which the information base for risk assessment and regulatory decision making might be improved. Several different approaches to such surveillance are possible. One approach might seek to augment or enhance spontaneous reporting. Another might build upon the infrastructure of other already established registries or databases. One focus of such approaches would be to provide a means of early case finding with the possibility of risk estimation. Another objective might be to provide ongoing feedback to regulatory authorities regarding the effect of regulatory interventions taken to improve the safety profile of a product.

In the realm of population based data resources, future developments could help to improve the information base available for regulatory risk management decision making. In too many instances, the only data available upon which to make risk assessments are those obtained from spontaneous case reports and drug use databases. Population based estimates of incidence are often lacking. The ADR of concern may be extremely rare or the drug product in question may have low levels of use, both indicating that the database lacks the statistical power to provide meaningful information. For new molecular entities marketed for less than three years, this is usually a severe limitation of existing population based resources. One solution would be for larger databases to be created and organized. It is possible that databases covering 10–20 million current lives might be required to permit study of rare or difficult to assess ADRs or those occurring with products new to the marketplace, with low levels of use in the

general population, or with use limited to a small segment of the population.

For most studies using data from longitudinal databases, access to medical records is important for purposes of confirming the diagnosis or outcome and for obtaining information on potential confounders and risk modifiers. Obtaining these records can be a difficult and time consuming process, and is becoming more difficult as concerns increase about confidentiality (see Chapters 19 and 26). Systems that rely on computerized medical records could substantially improve the efficiency of this aspect of study execution. Larger size and improved ease and access to primary records would improve the capacity of these databases to contribute valuable information on risk for a broad range of drug products, and within timeframes responsive to public health urgency and the need for timely risk management decisions to be made. The process of accessing medical record information should be done in a way that protects patient privacy but, at the same time, permits this critical function to occur in the interest of the public health.

Another important area of future development that could help to reduce the level of informational uncertainty centers on the concept of partnership. Many pharmaceutical companies are international in scope and many of the same pharmaceutical products are marketed in multiple countries throughout the world. Many of these countries have departments working at national centers dedicated to the task of postmarketing drug safety assessment. In this regard, there is a growing need for real-time collaboration between the various national centers and their postmarketing drug safety professionals. The challenge of drug safety and of the successful management of drug risk transcends national boundaries, as do the drug products under surveillance.

Similarly, there is growing need for partnership with industry in the task of monitoring for drug safety. Under ICH agreements, the Periodic Safety Update Report (PSUR) represents a potential step in that direction.⁵⁹ With the PSUR, pharmaceutical companies will assume greater responsibility for periodically conducting a comprehensive safety assessment of their products

and of submitting this to the regulatory authority of each country where the product is marketed. A PSUR is intended to provide an update of the worldwide safety experience of a medicinal product to competent authorities at defined times postauthorization. It should contain a critical evaluation of the benefit to risk balance of the product in the light of new or changing post-authorization information. This evaluation should also ascertain whether changes should be made to the marketing authorization, Summary of Product Characteristics (SPC), patient information leaflet, or product advertising.

According to European legislation, marketing authorization holders are currently required to submit a PSUR to competent authorities once a medicinal product is authorized in the EU, even if it is not marketed. PSURs are normally required to be submitted at six-monthly intervals for the first two years following the authorization, then annually for 2 years, and thereafter at each renewal (which takes place at 5 year intervals). Although the level of compliance to the system may still be increased, the quality of PSURs submitted by companies has gradually improved over the years and they are generally considered as a valuable source of information.

The concept of partnership can be extended to include other governmental agencies and organizations involved in the mission of public health as well as academia. In these ways, it may be possible to strengthen and improve the pharmacoepidemiologic safety net to the benefit of society.

REFERENCES

- Spilker B. *Guide to Clinical Trials*. New York: Raven, 1991; xx-xxiv.
- Rawlins MD, Jefferys DB. Study of United Kingdom product licence applications containing new active substances. *Br Med J* 1991; **302**: 223-5.
- Rogers AS. Adverse drug events: identification and attribution. *Drug Intell Clin Pharm* 1987; **21**: 915-20.
- Kessler DA. Introducing MedWATCH: a new approach to reporting medication and device adverse effects and product problems. *J Am Med Assoc* 1993; **269**: 2765-8.
- Gharaibeh MN, Greenberg HE, Waldman SA. Adverse drug reactions: a review. *Drug Inf J* 1998; **32**: 323-38.
- Pirmohamed M, Breckenridge AM, Kitteringham NR, Park BK. Adverse drug reactions. *Br Med J* 1998; **316**: 1295-8.
- Amery WK. Why there is a need for pharmacovigilance. *Pharmacoepidemiol Drug Safety* 1999; **8**: 61-64.
- Bakke OM, Manocchia M, de Abajo F, Kaitink I, Lasagna L. Drug safety discontinuation in the United Kingdom, the United States, and Spain from 1974 through 1993: a regulatory perspective. *Clin Pharmacol Ther* 1995; **58**: 105-17.
- Jefferys DB, Leakey D, Lewis JA, Payne S, Rawlins MD. New active substances authorised in the United Kingdom between 1972-1994. *Br J Clin Pharmacol* 1998; **45**: 151-6.
- Abenhaim L, Moore N, Bégaud B. The role of pharmacoepidemiology in pharmacovigilance. *Pharmacoepidemiol Drug Safety* 1999; **8**: S1-7.
- Rawlins MD, Thompson JW. Mechanisms of adverse drug reactions. In: Davies DM, ed., *Textbook of Adverse Drug Reactions*. New York: Oxford University Press, 1985; 12-38.
- Gillum JG, Israel DS, Polk RE. Pharmacokinetic drug interactions with antimicrobial agents. *Clin Pharmacokinet* 1993; **25**: 450-82.
- Richelson E. Pharmacokinetic drug interactions of new antidepressants: a review of the effects on the metabolism of other drugs. *Mayo Clin Proc* 1997; **72**: 835-47.
- Meyboom RHB, Egberts ACG, Edwards RI, Hekster YA, de Koning FHP, Gribnau WJ. Principles of signal detection in pharmacovigilance. *Drug Safety* 1997; **16**: 355-65.
- Wysowski DK, Baum C. Outpatient use of prescription sedative-hypnotic drugs in the United States, 1970 through 1989. *Arch Intern Med* 1991; **151**: 1779-83.
- Faich GA. Adverse reaction monitoring. *New Engl J Med* 1986; **314**: 1589-92.
- Griffin JP, Weber JCP. Voluntary systems of adverse reaction reporting: Part 2. *Adverse Drug Reaction Acute Poisoning Rev* 1986; **1**: 23-55.
- Olsson S. The role of the WHO programme on international drug monitoring in coordinating worldwide drug safety efforts. *Drug Safety* 1998; **19**: 1-10.
- Green G. Regional initiatives, past, present and future in the US. In: D'Arcy PF, Harron DWG, eds, *Proceedings of the Fourth International Conference on Harmonisation, Brussels, 1997*. Belfast: Queen's University, 1998; 425-9.
- Johnson JM, Barash D. A review of postmarketing adverse drug experience reporting requirements. *FDC Law J* 1991; **46**: 665-72.

21. Council for the International Organisation of Medical Sciences (CIOMS). *Final Report of the CIOMS Working Group II*. Geneva: CIOMS, 1992; 12.
22. Blum MD, Graham DJ, McCloskey CA. Temafloxacin syndrome: review of 95 cases. *Clin Infect Dis* 1994; **18**: 946–50.
23. Kubota K, Kubota N, Pearce GL, Inman WHW. ACE-inhibitor-induced cough, an unrecognized adverse reaction for several years: studies in Prescription Event Monitoring. *Eur J Clin Pharmacol* 1996; **49**: 431–7.
24. Khairi MR, Altman RD, DeRosa GP, Zimmermann J, Schenk RK, Johnston CC. Sodium etidronate in the treatment of Paget's disease of bone. A study of long term results. *Ann Intern Med* 1977; **87**: 653–63.
25. Hudson KF. A phenomenon of paradox: myocardial reperfusion injury. *Heart Lung* 1994; **23**: 384–93.
26. McKenzie R, Fried MW, Sallie R, Conjeevaram H, Di Besceglie AM, Park Y, et al. Hepatic failure and lactic acidosis due to fialuridine (FIAU), an investigational nucleoside analogue for chronic hepatitis B. *New Engl J Med* 1995; **333**: 1099–105.
27. Wechsler M, Drazen JM. Churg–Strauss syndrome. *Lancet* 1999; **353**: 1970.
28. Inman WHW, Weber JCP. The United Kingdom. In: Inman WHW, ed., *Monitoring for Drug Safety*, 2nd edn. Lancaster: MTP Press, 1986; 13–47.
29. Scott HD, Rosenbaum SE, Waters WJ, Colt AM, Andrews LG, Juergens JP, et al. Rhode Island physicians' recognition and reporting of adverse drug reactions. *RI Med J* 1987; **70**: 311–6.
30. Rogers AS, Israel E, Smith CR. Physician knowledge, attitudes and behavior related to reporting adverse drug events. *Arch Intern Med* 1988; **148**: 1596–1600.
31. Inman WHW, Adelstein AM. Rise and fall of asthma mortality in England and Wales in relation to use of pressurized aerosols. *Lancet* 1969; **2**: 279–85.
32. Belton KJ, Lewis SC, Payne S, Rawlins MD, Wood SM. Attitudinal survey of adverse drug reaction reporting in the United Kingdom. *Br J Clin Pharmacol* 1995; **39**: 223–6.
33. Eland IA, Belton KJ, van Grootheest AC, Meiners AP, Rawlins MD, Stricker BHCh. Attitudinal survey of voluntary reporting of adverse drug reactions. *Br J Clin Pharmacol* 1999; **48**: 623–7.
34. Kurz X., Van Ermen A., Roisin T, Belton KJ. Knowledge and practice of adverse drug reaction reporting by Belgian physicians. *Arch Public Health* 1996; **54**: 29–41.
35. Biriell C, Edwards IR. Reasons for reporting adverse drug reactions—some thoughts based on an international review. *Pharmacoepidemiol Drug Safety* 1997; **6**: 21–26.
36. Wood SM, Waller PC. Record linkage databases for pharmacovigilance: a UK perspective. In: Walker SR, ed., *Databases for Pharmacovigilance*. Carshalton: Centre for Medicines Research, 1996; 47–54.
37. Department of Health and Human Services. Studies of adverse effects of marketed drugs; availability of grants (cooperative agreements); request for application. *Fed Reg* 1998; **63**: 33680–6.
38. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *Br Med J* 1991; **302**: 766–8.
39. Thompson D, Oster G. Use of terfenadine and contraindicated drugs. *J Am Med Assoc* 1996; **275**: 1339–41.
40. Ray WA, Griffin MR, Downey W. Benzodiazepines of long and short elimination half-life and the risk of hip fracture. *J Am Med Assoc* 1989; **262**: 3303–7.
41. Garcia Rodriguez LA, Jick H. Risk of gynecomastia associated with cimetidine, omeprazole, and other antiulcer drugs. *Br Med J* 1994; **308**: 503–6.
42. Hirschorn LR, Trnka Y, Onderdonk KA, Lee ML, Platt R. Epidemiology of community-acquired *Clostridium difficile*-associated diarrhea. *J Infect Dis* 1994; **169**: 127–33.
43. Graham DJ. Epidemiologic review of spontaneous case reports of withdrawal seizures, physiologic dependence and death reported with alprazolam, with special attention to effects of high dose and longer duration use. Psychopharmacologic Drugs Advisory Committee Meeting, Food and Drug Administration, Rockville, MD, 1989.
44. Stockley IH. *Drug Interactions*. London: Pharmaceutical Press, 1996; 1–3.
45. Garcia Rodriguez LA, Stricker BH, Zimmerman HJ. Risk of liver injury associated with the combination of amoxicillin and clavulanic acid. *Arch Intern Med* 1996; **156**: 1327–32.
46. Begaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? *Br J Clin Pharmacol* 1994; **38**: 401–4.
47. Shaw FE, Graham DJ, Guess HA, Milstein JB, Johnson JM, Schatz GC, et al. Postmarketing surveillance for neurologic adverse events reported after hepatitis B vaccination: experience of the first three years. *Am J Epidemiol* 1988; **127**: 337–52.
48. Nightingale SL. Recommendation to immediately withdraw patients from treatment with felbamate. *J Am Med Assoc* 1994; **272**: 995.
49. Duract (bromfenac) “Dear Healthcare Professional” letter. Wyeth–Ayerst Laboratories. 22 June 1998.
50. Moore N, Kreft-Jais C, Haramburu F, Noblet C, Andrejak M, Ollagnier M, et al. Reports of hypoglycemia associated with the use of ACE inhibitors and other drugs: a case—not case study

- in the French Pharmacovigilance System database. *Br J Clin Pharmacol* 1997; **44**: 513–8.
51. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA Spontaneous Reporting System. *Am Stat* 1999; **53**: 177–90.
 52. 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.
 53. Stang MR, Wysowski DK, Butler-Jones D. Incidence of lactic acidosis in metformin users. *Diabetes Care* 1999; **22**: 925–7.
 54. FitzSimmons SC, Burkhart GA, Borowitz D, Grand RJ, Hammerstrom T, Durie PR, *et al.* High-dose pancreatic-enzyme supplements and fibrosing colonopathy in children with cystic fibrosis. *New Engl J Med* 1997; **336**: 1283–9.
 55. Smyth RL, Ashby D, O’Hea U, Burrows E, Lewis P, van Velzen D, *et al.* Fibrosing colonopathy in cystic fibrosis: results of a nested case–control study. *Lancet* 1995; **346**: 1247–51.
 56. Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. *J Am Med Assoc* 1991; **266**: 2847–51.
 57. Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, *et al.* for the ADE Prevention Study Group. Systems analysis of adverse drug events. *J Am Med Assoc* 1995; **274**: 35–43.
 58. Lee A, Bateman N, Edwards C, Smith JM, Rawlings MD. Reporting of adverse drug reactions by hospital pharmacists: pilot scheme. *Br Med J* 1997; **315**: 519.
 59. International Conference on Harmonisation; guidelines on clinical safety data management: Periodic Update Reports for marketed drugs; availability. *Fed Reg* 1997; **62**: 27470–6.

9

A View From a US Courtroom

JAMES R. RUGER* and SIDNEY H. WILLIG

Sanofi, Inc. Legal Department, New York, USA; *and Siemens Corporation, Iselin, New Jersey, USA

INTRODUCTION

Pharmacoepidemiologic evidence may be admitted to the courtroom to help prove legal issues such as failure to warn, or to demonstrate the presence or absence of valid associations between an event (i.e., the use of a drug) and an outcome (i.e., an adverse drug reaction). Due to the complex nature of epidemiologic studies, the use of these data has many times led to inappropriate interpretations in the courtroom.^{1,2} The situation worsens when “experts” lacking training in epidemiology are allowed to testify, because their testimony may be erroneously persuasive to a lay jury and, thus, misunderstood.

For all of these reasons, the importance of pharmacoepidemiology to the legal process, and the importance of its proper use, is clear. This chapter will explore this role in more detail. It will begin with some general background information on drug product liability litigation and the burden of proof it presents for each side. The differences among medical causality, epidemiologic causality, and legal causation will follow. Then a description will be provided of the impact of adverse drug event (ADE) reporting on litigation. Finally, a series of specific examples of these principles will be provided. While the viewpoint taken here is that

from the US courtroom, pharmacoepidemiology litigation is most widespread in the US, so it is perhaps the best model to examine. Regardless, US litigation affects drug manufacturers, regulators, and academics around the world, as well as worldwide drug availability, and sometimes serves as a model, which is perhaps adopted too frequently elsewhere. For all of these reasons, the US perspective on litigation should be of interest to pharmacoepidemiologists around the world.

REGULATION VERSUS LITIGATION

THE FOOD AND DRUG ADMINISTRATION'S ROLE

With statutory authority from the Federal Food, Drug, and Cosmetic Act (as amended by the FDA Modernization Act of 1997), the FDA's role in regulating pharmaceutical manufacturers has included evaluating evidence of efficacy and safety,^{3,4} deciding on the approvability of new drug applications,⁵ deciding on the withdrawal of new drug applications,⁶ and recording all relevant information concerning the effectiveness and safety of marketed drugs.⁷ To accomplish these goals, certain implementing regulations have been

promulgated⁸ and guidelines issued. These guidelines state procedures or standards of general applicability that are not legal requirements, but are procedures acceptable to FDA for a subject matter that falls within the laws it administers. The vastness of the FDA's task is then enhanced and assisted by a regulatory and administrative framework that solicits and, in fact, requires reports by manufacturers of any adverse experiences observed during administration of the drug in any manner (see also Chapter 10). This, in turn, generates a controversy over the best methodology for assessing causality in case reports (see Chapter 32).

Thus, the Federal Food, Drug, and Cosmetic Act, legislated as a public safeguard, seeks to maintain the quality of the products over which it has governance.⁹ It even offers economic protection against adulteration and misrepresentation. However, it is not restitutive.¹⁰ It does not reimburse those who have been harmed by adverse drug effects. Rather, it relies on the seizure and condemnation of noncompliant products and injunctive and criminal sanctions against wrongdoers.¹¹ Therefore, when individuals seek redress for an injury from a drug, vaccine, cosmetic, or device, they need to use the usual pathways for establishing civil liability, that is litigation.

The great importance of the statutory direction of the federal law is that it has established a duty for the Food and Drug Administration to monitor the manufacturing and distribution of drug products. The essence of that duty is, however, shared by those who research, develop, manufacture, market, and distribute those drugs. Thus, whereas in legal understanding, one who makes a claim for alleged injury must establish the existence of a legal duty owed by the defendant, here the law and regulations point toward such a conclusion. In addition, as will be explained in more detail below, failure to follow the law can make an important contribution to a product liability case.

Pursuant to requirements of the FDA Modernization Act of 1997 (FDAMA), section 555, if the FDA receives new information indicating that a new use of a drug may not be effective or may present a significant risk to public health, the agency may take appropriate action, including

ordering the manufacturer to cease dissemination. It is certainly conceivable that such data could be of a pharmacoepidemiologic nature. Further, section 113 of FDAMA adds a new subsection (j) to the Public Health Service Act (PHSA) (42 U.S.C. section 282) that requires the Secretary of Health and Human Services to establish a coordinated program within the National Institutes of Health to create, maintain, and operate a data bank of information on clinical trials for drugs for serious or life threatening diseases and conditions. This data bank may contain information on the results of clinical trials, including information on potential toxicities. Thus, the data bank may provide source information for pharmacoepidemiologic followup.

PLAINTIFFS' BURDEN OF PROOF IN PRODUCT LIABILITY CASES

Judgments for the plaintiff in product liability litigation are usually based on either of two legal theories: negligence or strict liability. Rarely, contract theory (via breach of warranty) or more extreme tort theory (such as fraud) comes into play.

For judgments based on showing negligence or fault, a plaintiff needs to show that there was a duty that was owed to him. He or she also needs to show that there was a dereliction from such duty that directly caused his economic and/or personal injury. Finally, he or she needs to show that this injury resulted in damages that are measurable and can be redressed by monetary award. For these actions, the court and the jury are looking primarily at whether there was defective conduct by the defendant(s).

For judgments based on strict liability, a plaintiff must show that a product was defective in some respect, whether due to defective design, construction, instructions for safe and effective use, or warnings. Further, to obtain a judgment based on strict liability, a plaintiff must show that the product came from the supplier in that defective condition, making it unreasonably dangerous, and caused the economic and/or personal injury that is the subject of the complaint. Thus, the defendant's state of mind—his negligence, his

fault—is not an issue. The defective or nondefective condition of his product is the sole issue. The connection between conduct and product defect is then made only by plaintiffs seeking additional punitive or exemplary damages. The ultimate question that then arises is whether it is always defective conduct to sell a bad product.

Regardless of whether the legal theory is negligence or strict liability, expert testimony is required on the issue of causation, i.e., whether the product caused the injury in question. The sophistication of the pharmacology of prescription drugs, vaccines, and antibiotics precludes a jury from drawing an inference of causation based upon common knowledge; expert testimony is critically important. In fact, a jury could not be given a *res ipsa loquitur* instruction (the thing speaks for itself) based on its common knowledge. A plaintiff is not precluded from setting forth the argument that he did not have polio, he ingested polio vaccine, and then he had polio, so the vaccine must have caused the polio, or at least it failed to prevent the polio it was supposed to prevent. The jury, of course, may or may not accept that argument.

In addition, if a plaintiff expects to prove that a product was defective due to a deficient warning, his or her burden of proof is somewhat similar, regardless of whether his argument is couched in strict liability or negligence terms. A plaintiff may attempt to establish a *prima facie* case based on evidence that the drug functioned improperly, but he is still not relieved of the burden of proving the product was in defective condition when it came to his use. The *res ipsa* approach in medical malpractice litigation,¹² and in product liability litigation to a lesser extent (where it applies to negligence rather than strict liability theory), also requires a plaintiff to show that the product failure, more probably than not, was due to the defendant. The plaintiff must show that the resultant injury was not one that could have happened if it were not for the defendant's negligence, and no explanation is available other than defendant's negligence. The plaintiff says "I did nothing to create my injury, nor did anyone prescribing, administering or applying the drug to me."¹³ He is, however, permitted to prove the

existence of a defective condition by offering circumstantial evidence. Circumstantial evidence is that which translates recognized human experience into a credible slant on events. For example, ordinarily one does not suffer an automobile tire blowout after only four miles of use unless the tire had a defect when received. Bolstered by additional facts, whether direct evidence or even more circumstantial evidence from another source, this type of evidence may be offered as proof of a defect. Conceivably, pharmacoepidemiologic testimony could give some weight to any circumstantial evidence that might be offered to make or refute a claim of injury associated with the use of a drug product, by indicating that it was plausible that the drug could have caused that injury.

Compounding the difficulty in connecting ADEs with causation is the nature of drug product liability suits today. In product liability suits based upon negligence, proximate cause involves apportionment of responsibility for an injury, and apportionment of damages upon perhaps several causes. In terms of strict drug product liability, however, whatever the product defect claimed—whether as to design, instructions for administration, or warnings on use—it must be shown to be the direct cause of the injury claimed. What if the proof offered creates equal opportunity for more than one direct cause? Can there be more than one direct cause? What if one cause represents a possibility and no more? What if one cause represents a possibility and another represents a probability? These "what ifs" have definite answers in law, but they tend to be answered by triers of fact—the jury—rather than by a judge, who rules as to law.

Can a strict liability demonstration of a product defect be based on possibility? As an individual question, that is addressed in every lawsuit in which the plaintiff has *only* pharmacoepidemiologic evidence to offer, alone and via the testimony of medical experts. The responses are not necessarily consistent. Some attorneys feel that the personalities of the adversaries and their counsel, the reputation of those who testify, the venue, and jurisdictional factors have a profound influence on outcome in such cases.

For example, federal judges have, in recent years, relied heavily on the expertise possessed by federal agencies such as the FDA. The state trial courts, however, are much less apt to feel the same about FDA expertise. For them, most times federal agency approval of products, procedures, labeling, or compliance with federal regulations are not necessary proof of rectitude or even good intentions. Thus, the fact that the FDA has complete records and reports concerning ADEs reported following use of a drug product that remains on the market does not of itself imply that the labeling or design of the product are non-defective. Yet, all too often state courts seem taken in by the “false aura of scientific infallibility, coupled with low probative value” that are imparted by analysis of any number of spontaneously generated ADEs by a party’s pharmacoepidemiologic witness in a lawsuit.

This theory is supported by the Richardson case,¹⁴ since a virtually companion case which was tried in the same jurisdiction earlier, but in a nonfederal court system,¹⁵ resulted in a substantial plaintiff’s judgment rather than an overturned judgment. As District Judge Jackson, in the federal case, described it,

This trial was a virtual reprise of “Oxendine” ... The trial judge ... set aside the verdict ... but the district of Columbia Court of Appeals, reversed an opinion ... bearing upon the current letter of scientific knowledge. It reopened a 20-year-old controversy which is by now essentially settled within the scientific community. Reasonable jurors could not reject scientific consensus that Bendectin® is not a teratogen without indulging in speculation and conjecture.”¹⁴

DRUG PRODUCT LIABILITY DEFENSE

A defendant, in turn, will claim that when the drug product left his possession, it was not in defective condition. He may further attempt to show that it was used improperly or abnormally used. He also may introduce evidence about any reasonable secondary causes that may eliminate him from liability, as the initial burden of proof is on the plaintiff to show the product defect or the negligent conduct.

The two bellwethers of drug products liability defense have been, with regard to prescription drugs and devices, the “learned intermediary rule” and the “state of the art” defense.

The learned intermediary rule, while still accepted in most jurisdictions, may be eroding. In those locales where the learned intermediary rule still exists (the majority of states), it is the prescribing health care professional to whom the manufacturer’s duty to warn runs, not the patient.¹⁶ However, the implementation of direct to consumer advertising and the advent of the internet have created new issues not existing previously in the prescription drug arena. If a drug company chooses to unduly influence the public regarding a drug product and fails to warn properly, the learned intermediary defense may be lost. Thus, the special circumstances as previously set forth in *Reyes v. Wyeth*¹⁷ and *Davis v. Wyeth*¹⁸ and their progeny in the vaccine area and *MacDonald* in the contraceptive field¹⁹ may no longer be the exception to the rule, since it may be argued that, via direct to consumer advertising and internet promotion, the traditional relation between physician and patient is lost. Thus, due to the intentional acts of pharmaceutical manufacturers, a key defense may be forfeited. Additionally, for the “learned intermediary rule” to serve its function as prophylactic against liability, the package insert accompanying the product must have been adequately and honestly revelatory. Otherwise, as the patient’s agent, the prescriber could not weigh accurately the benefits against the risks.

The state of the art defense is evolving as well. “State of the art,” with respect to prescription drugs, has represented the quality of warning and instruction, based on the best information the product labeler could and should have available. These warnings are intended to give the prescribing or administering physician an adequate basis for making a benefit to risk assessment in the interests of the patient. This has been discarded by some jurisdictions with respect to problems other than these.²⁰ However, this too is changing, due to the finalization of the Restatement 3rd of Torts. Many jurisdictions still adhere to the Restatement 2nd. As time passes, however, the newer version

will most likely become the rule. The Restatement 3rd holds liable anyone who sells a prescription product, unless the learned intermediary rule is in existence. It also diminishes some of the impact of comment (k) to section 402 of the Restatement 2nd concerning design defect. Under Restatement 3rd, one may be liable for a design defect if a prescription drug is found lacking efficacy in all instances. Thus, the burden would be to prove either the FDA made an erroneous decision, or it is totally irrational for the drug to be used under all circumstances. This is a tough burden to meet. It is still recognized that pharmaceutical products are unavoidably unsafe, as pointed out in comment (k) to section 402 of Restatement 2nd.²¹ The requirement also remains that the manufacturer needs to supply all the information known about the product's safe and effective use that should reasonably be known to any expert who would be producing the product at that time.

THE IMPORTANCE OF COMPLIANCE WITH THE STATUTE TO PRODUCT LIABILITY LITIGATION

When an injury is accompanied by a statutory breach, the party injured need not necessarily be required to prove common law proximate causation.²² However, he or she must at least show that the claimed injuries resulted in whole or in part from violation of the statute. In the cited case,²³ the court held that a violation of the Safety Appliance Act rendered carriers absolutely liable "for injuries resulting therefrom, without regard to negligence or time at which the malfunction occurred." In a similar and unrelated lawsuit, citing the Federal Employers Liability Act,²⁴ the court said that some showing of negligence would be required of plaintiff. However, "... any showing of negligence on the part of the railroad, even the slightest, is sufficient to create liability," so the employer can suffer liability for negligence that does not rise to the level of common law negligence.²⁵ Further, one can readily look at more recent litigation, where defendants' failures to meet statutory and regulatory requirements fully and truthfully occasioned billions of dollars in adverse judgments and even corporate "wipe-outs." These

include the "MER-29" cases,²⁶ the "Chloromycetin" cases,²⁷ the "Selacryn" cases, the "Oraflex" cases, the Dalkon Shield cases, the "toxic shock" tampon cases,²⁸ etc.

Thus, there is no denying now that statutes, and regulations promulgated thereunder, create duties whose breach feeds liability for the individual or business entity subject to that regulation. Certainly the failure to implement FDA requirements cost defendants heavily in *Stanton v. Astra*, where the reporting requirements about ADEs were declared by the plaintiffs to have been flouted.²⁹ In that instance, the manufacturer, confident its anesthetic was no longer to be considered a new drug, did not deem itself required to file ADEs. The jury believed the plaintiff's claim that, had they done so, the resulting warnings would have been believed and would have spared the injury to the child (see below).

However, when breach of statute is proffered by a plaintiff in making the so-called "negligence *per se*" position, the trial judge must undertake the responsibility for informing the "jury, in substance, of any possible excuse or justification, which might be applicable to the case, and how such excuse or justification would affect the parties" respective burdens of proof.²² There is little question that the substance of such an instruction must be introduced by the attorneys as part of their case, however. Judges cannot be expected to have sufficient independent understanding of what might represent possible excuses and justifications in drug, vaccine, or medical device product liability.

As an example, in *Stanton v. Astra*, mentioned in an earlier example,²⁹ the judge found no possible excuse or justification for the absence of ADE reporting by Astra. Yet, there had been no judgment countering Astra's position that "old" drugs according to the statutory "nondefinition" were not required to file ADEs. In fact, it was only in 1986 that the FDA formally declared this, by regulations published in the Federal Register and included in the Code of Federal Regulations. The FDA operates on a principle that a drug never leaves new drug status until they say so, yet the manufacturer's attorneys argued that the statute clearly indicates when a drug is no longer a new

drug. The FDA took no action against Astra for failure to file ADEs for Xylocaine[®], although it clearly indicated its position in favor of the plaintiff at trial by testimony of FDA staff. The question still remains, therefore, whether the defendants could have urged the judge to explain to the jury that, regardless of other proofs mustered, the presumption of negligence created by the claimed breaks of statute and regulations could stand rebutted if Astra showed it acted as a reasonably prudent person would have acted under the same circumstances.

Within the considerations for court and jury will be the facts and evidence mustered by both parties on causation of the injury. Proximate cause of injury always seems to involve disputed questions of fact. Where statutory breach or regulatory noncompliance is sought to be equated with proximate cause of the plaintiff's harm, while the burden of proof ostensibly remains with the plaintiff, he at least seemingly enjoys the presumptive advantage flowing from the "negligence *per se*" theory. Most courts will agree that the best rule in negligence alleging statutory or regulatory violation is that the plaintiff must first prove the violation, and that of itself will create a rebuttable presumption of negligence. However, even assuming a violation of the statutory duty spelled out in the Act, a question of fact will remain in existence as to whether that violation is the proximate cause of the injury. Therefore, given appropriate instruction on the point, and examining the evidence before it, a jury is free to find that a statutory breach is not necessarily the direct cause of the injury.³⁰ There is always the possibility that conflicting facts and inferences relevant to the issue of proximate cause exist. As an example, in the host of decisions adverse to Parke-Davis in the Chloromycetin[®] cases, there was an underlying statutory breach involved with misbranding and new drug violations, according to the Federal Food, Drug, and Cosmetic Act. Yet, where the treating physician averred that he did not read the labeling and so was not influenced by such violations, but had prescribed the drug solely on the basis of his own independent knowledge and judgment, the Parke-Davis statutory and regulatory breaches failed to amount to proximate cause.²⁷

CAUSALITY ASSESSMENT

MEDICAL VERSUS LEGAL VERSUS EPIDEMIOLOGIC CAUSALITY

If pharmacoepidemiologic findings, as expressed by experts, are to be presented to a court in support of a theory of injury, causation by a particular drug becomes a battleground between different experts. Observers, assessors, and evaluators with access to the same body of information about the drug and the adverse event associated with it should seemingly find consistent threads of causation and noncausation. But, as we examine the tonnage of "swine-flu" and Bendectin[®] testimony, to name just a few instances, along with Judge Weinstein's findings in the "Agent Orange" cases,³¹ we find that this is not so. Must the answer lie in the expert's elevation of his own subjective deliberations as overshadowing whatever objective sources may exist for the given ADE? Determining medical causation in individual cases is discussed more in Chapter 32. However, as Hutchinson says, "The standard adopted in the causality assessment literature is the unaided judgment of experts, supplemented by some consensus-producing convention when the experts disagree with one another".³²

For litigation, such a medical causality assessment, even consensually produced, may not meet the criteria for legal causation, i.e., proximate cause. The legal view of causation, because it was developed originally for almost exclusively economic redress, places heavy emphasis upon the proof necessary for one seeking such redress. "Actual cause has often been confused with proximate causation. The significance of proximate cause focuses upon legal policy in terms of whatever responsibility will be extended to the consequences of conduct which has occurred. Actual cause, however, is a factual question focusing on the antecedent factors producing a particular consequence".³³ The medical view of causation permits exploration and hypothecation based upon a high degree of subjective analysis brought to bear on whatever objective data are available. It allows for one to propound a theory of causation on first impressions. It gives full or

some consideration to any and all hearsay evidence. A physician who has never seen a case of periarteritis nodosa, as a medical student or a practitioner, nor has practiced in a medical specialty related to such a disease entity, can examine a patient's record, find some criteria that resemble those of the disease, and unhesitatingly testify that the plaintiff suffers from that disease and contracted it from exposure to the particular drug. That is not to say that he will not be rebutted by another expert, but it does say that he has made what is, for him, a satisfactory causality assessment and will testify to it under oath. It is a judgment a jury may not find deficient.² Legal writers stress "the principle that a logical, although not necessarily an empirical, fallacy lies in *post hoc, ergo propter hoc* reasoning".³⁴ Defendants complain that courts are too ready to recognize a close sequence between the defect alleged and/or proven and causation. That is understandable in light of medical causality assessment, which never has available the formal and diverse rendition of facts and expert testimony mustered at a trial.³⁵

Thus, there is a major difference between medical causality assessment and legal causation, one which should be, but often is not, appreciated by attorneys, courts, and expert witnesses. A spontaneous reporting system, with its individual medical causality assessment, deals in possibilities. Yet in the law, "possibility" is modified by the adjective "mere." As Cardozo stated in his classic adaptation of our common law to the 20th century tort system, one does not find liability for negligence on mere possibility. That requires probability. "There must be knowledge of a danger, not merely possible, but probable. It is possible to use almost anything in a way that will make it dangerous if defective. That is not enough to charge a manufacturer with a duty independent of his contract".³⁶ Minus probability, the great Justice said, there could be no element of foreseeability, no use of the reasoning process in formulating warning or desisting from an action.³⁷ Man can only be reasonable in his exercise of foresight when he makes his assessment on that basis.

Epidemiological analysis, on the other hand, biostatistically creates a basis for predictability.

This has been recognized by the courts. Where a plaintiff's decedent had worked in a uranium mine and died of cancer, the cause was supplied from epidemiologic study and testimony virtually alone.³⁸ However, as this author previously pointed out, epidemiologic evidence, if used to assert and prove causation, "must meet the qualifications that traditionally distinguish real evidence. Like statistical studies, they may offer bricks to build," or to fortify such theory of causation.³⁹

The process of deciding about whether or not there is epidemiologic causality is discussed in detail in Chapter 2. It is based on finding a statistical association, and then using established criteria to judge whether the finding is a causal one. Importantly, however, epidemiology uses inductive reasoning, not the deductive reasoning that is far more familiar to the attorney who employs an epidemiologist as an expert. Thus, an epidemiologic statement about causality generalizes from individuals to groups and populations, not the reverse. Unless the disease in question occurs very rarely in the absence of the exposure, epidemiology is unable to determine the cause in a particular case. Although epidemiology represents a systematic and a rigorous basis for deciding on cause, it still will not generally get one to a legal cause. This requires a medical judgment, although this in turn often needs to be based on rigorous epidemiologic data.

However, the results of thorough and validated pharmacoepidemiologic studies are far more often determinative than the results of testing animals or animal tissue. At best, as was pointed out in "Agent Orange," citing a Dr Silbergeld, "animal studies are aimed at discovering a dose-response relationship while epidemiologic studies show an association between exposure and disease".⁴⁰ Time may give us additional assurance that the use of pharmacoepidemiologic methods will prove to be a most valuable means toward more complete, more valid, and more contemporary information about the effects of the potent and widely used drugs of this period. Time may also provide a better perspective on the way the courts should receive evidence, testimony, and exhibits based upon the reports and records that created the basis for

pharmacoepidemiologic analyses and statements of causation.

IMPACT OF PHARMACOEPIDEMOLOGY ON LITIGATION

EXPERT TESTIMONY AND THE COURTS

The potential usefulness of epidemiologic evidence in helping to establish associations between exposure to physical and chemical substances, including pharmaceutical preparations, and injuries alleged in products liability litigation has been recognized for a number of years. Scrutiny by courts as to the admissibility of expert testimony to assist in interpreting epidemiologic data and opining thereon initially assumed special meaning in a number of cases involving an area of law known as toxic torts. The common thread among these litigations was the allegation of a latent injury (i.e., injury that manifests years after exposure). Asbestos, Agent Orange, dioxin, radiation, polychlorinated biphenyls (PCBs), and diethylstilbestrol (DES) are prime examples. Epidemiology has been used in these matters in an attempt to demonstrate associations between exposure to the substance in question and an increased risk to the exposed population of developing certain diseases. The cases discussed in this section, both pharmaceutical and nonpharmaceutical, have established the legal precedents regarding epidemiology and its relation to the law.

The use of epidemiology in litigation soon expanded beyond latent injury claims (e.g., cancer) to claims involving birth defects (e.g., Bendectin®) and other more immediate injuries (e.g., vaccines and neurologic injury, aspirin and Reye's Syndrome, xenobiotics and immune dysfunction, etc.). In many of the birth defect cases, a path was traced from administration or ingestion of an agent by the mother, to the fetus or newborn, and then to the resultant childhood, adolescent, or adult physical or mental damage. Epidemiologic evidence has then been used to help establish or disprove (via studies showing a lack of correlation

between exposure and defect) an increased risk to the offspring.

In order to appreciate the evolution of how courts have dealt with epidemiologic evidence and the experts who testify concerning their interpretations of that evidence, a broader range of cases than exists solely within the pharmaceutical sphere must be discussed. While other sections of this chapter deal primarily with pharmaceutical cases, this section will discuss both pharmaceutical and nonpharmaceutical substances that have been the subject of litigation throughout the years.

Some of the earliest cases that considered these issues entailed close scrutiny and active involvement by the courts in determining the admissibility of expert testimony and interpreting the significance of both the data relied upon by expert witnesses and their credentials. *Johnston v. The United States* is illustrative.⁴¹ *Johnston* involved an action by four employees of an aircraft instrument and development plant who brought suit against the United States Government, alleging that their various cancers were caused by exposure to radiation while at their jobs. The court scrupulously analyzed the epidemiologic and scientific data as well as the experts presented by both sides. It determined that many of the plaintiffs' experts were either advocates, alarmists, or that they were relying on unreliable data to support their opinions. The court also analyzed statistical risk calculations presented by the experts and determined that some of plaintiffs' experts were using statistics in a speculative way. In its opinion, the court noted that an expert can easily manipulate numbers to, "... have a final result of over 50%, i.e., more probable than not." Thus, plaintiffs' case was dismissed, even though their experts opined that the plaintiffs' cancers were "more likely than not" caused by exposure to workplace radiation.

Allen v. The United States provides another early example of the trial court evaluating scientific data proffered as evidence by experts.⁴² The *Allen* court, after discussing the difficulty that exists in establishing causation where latent, toxic injuries are concerned, broached the concept of statistical relationships, especially those arising from epidemiologic evidence. The court examined each

individually named plaintiff's case, specifically looking at the following factors:

1. whether the plaintiff with cancer was probably exposed to radiation significantly in excess of background radiation rates;
2. whether the injury alleged is consistent with those known to be caused by ionizing radiation; and
3. whether the person injured resided in the geographic area in question in the lawsuit.

The court decided that, if these points were established, other relevant factors should then be evaluated in order to determine whether the "substantial factor" test was met by the claiming plaintiff. Epidemiologic data were used by the court to assess a number of these parameters, as well as the reported latency periods for various cancers. Upon completion of its analysis, the court determined that ten of the 24 representative cases warranted compensation.

As is discussed in other sections of this chapter, Judge Weinstein, in granting the defendants summary judgment regarding those plaintiffs who opted out of the class action in *in re: "Agent Orange"*, analyzed the deposition testimony and affidavits of plaintiffs' experts.⁴³ He found that, pursuant to the Federal Rules of Evidence, their opinions were inadmissible. Judge Weinstein determined that both the opinions of expert witnesses and the basis for their opinions should be evaluated by the court and that the Federal Rules of Evidence permit such an evaluation. Federal Rule of Evidence 702 states that an expert who is qualified by knowledge, skill, training, or education may testify in the form of an opinion as to technical, scientific, or other specialized knowledge if it will assist the trier of fact (generally a jury) in understanding the evidence. Judge Weinstein used this rule in conjunction with Federal Rule 403 (which allows a judge to exclude relevant evidence if it has a propensity to unfairly prejudice, confuse, or mislead a jury), to exclude the proposed testimony of a number of plaintiffs' experts. Further, he invoked Rule 703 to show that the facts or data upon which plaintiffs' experts were relying were not of the type upon which experts in

the particular field of expertise would reasonably rely. On appeal, Judge Weinstein's decision was upheld by the Second Circuit Court of Appeals.

In *Viterbo v. The Dow Chemical Co.*, the trial court dismissed plaintiff's allegation that the herbicide, picloram, caused his mental depression.⁴⁴ The court examined the credentials of plaintiff's experts, their methodologies, and the basis upon which they formed their opinions. It determined that the data they used to form their opinions were unreliable and not of the type reasonably relied upon by experts in the field. Their testimony was ruled inadmissible, pursuant to the Federal Rules of Evidence Rule 703. The *Viterbo* decision was appealed to the Federal Court of Appeals.⁴⁴ In rendering its opinion affirming the trial court, the Court of Appeals wrote, "In this case today we consider the question whether it is so if an expert says it is so." The answer from the Court was a resounding, "no". The Fifth Circuit reasoned that, although Rule 703 expanded what was the acceptable basis of expert opinion in common law, this expansion does not make summary judgment impossible whenever a party produces an expert to support its position. The court went on to state that:

As a general rule, questions relating to the bases and sources of an expert's opinion affect the weight to be assigned that opinion rather than its admissibility and should be left for the jury's consideration In some cases, however, the source upon which an expert's opinion relies is of such little weight that the jury should not be permitted to receive that opinion. Expert opinion testimony falls into this category when that testimony would not actually assist the jury in arriving at an intelligent and sound verdict.

Numerous other cases have followed the reasoning espoused in the above cases, e.g., various Bendectin[®] cases, described elsewhere;^{14,15} *Thomas v. Hoffman-LaRoche, Inc.*;⁴⁵ *Novak v. United States*;⁴⁶ *Sterling v. Velsicol Chemical Co.*;⁴⁷ and *Boyles v. American Cyanamid Co.*⁴⁸

However, there are also a number of cases that stand for the opposite proposition: that the trial court should not preclude expert testimony if the expert is qualified and uses methodologies

commonly accepted by others in the field. These cases hold that it should not be the expert's opinion itself that a court should evaluate but, rather, the court should look at the methods used by the expert in arriving at an opinion and whether the methodology used is one that is generally accepted by the relevant scientific community. The court should allow the jury to decide the veracity of the testimony and of the opinion itself. An example is *Ferebee v. Chevron Chemical Co.*,⁴⁹ which was brought by the children of an agricultural worker who died of pulmonary fibrosis against a defendant that manufactured the herbicide paraquat.

After the jury found for the plaintiffs, the defendant appealed on the grounds that the jury should never have been allowed to consider the testimony of plaintiffs' experts because their expressed opinions and theories were not generally accepted in the scientific community. In rendering its opinion, the appellate court stated that, even though an expert's opinion may be controversial in its conclusions, it should be admitted if it is based upon wellfounded methodologies. The court deferred to the jury in weighing the expert testimony and emphasized that judges, both trial and appellate, have no special competency to resolve complex causal issues such as arise in toxic tort cases.

In the *Paoli Railroad Yard PCB Litigation*,⁵⁰ the trial court dismissed plaintiffs' claims based upon a finding that plaintiffs' expert witnesses were espousing "junk science" regarding the alleged health effects of PCB exposure. The court reviewed the literature on PCBs and determined that, "... no reasonable or scientific basis exists for concluding that chronic exposure to PCBs causes cancer, hypertension, cardiovascular diseases, elevations in serum triglycerides or cholesterol levels, liver disease, joint irritation, pancytopenia ...". The trial court in *Paoli* (1989) stressed that "A vigorous examination is especially important in the toxic tort context, where presentation to the trier of theories of causation depends almost entirely upon expert testimony." The US Court of Appeals for the Third Circuit reversed this decision (1990). The Third Circuit adopted a good part of the reasoning used in *Ferebee v. Chevron*. It

concluded that almost all of the plaintiffs' expert opinions were improperly excluded because of a misapplication by the trial court of the test for "novel scientific theories".⁵⁰

On remand back to the trial court, six days of hearings were then held concerning defendants' motion to exclude animal studies, certain laboratory analyses, and the testimony of ten of plaintiffs' experts, pursuant to Federal Rules of Evidence 403, 702, and 703. The trial court did just that. As a result, the case is still pending.

*Peteet v. The Dow Chemical Co.*⁵¹ is another example where a novel scientific theory was allowed to stand despite its lack of general acceptance in the scientific community. This case was originally part of the Agent Orange multi-district litigation but was severed therefrom because it did not involve military exposure to Agent Orange. As a twist, in *in re "Agent Orange"* the herbicidal component in question was 2,4,5-T (2,4,5-trichlorophenoxy acetic acid) which was allegedly contaminated with dioxin (2,3,7,8-tetrachlorodibenzo-*p*-dioxin). Yet in *Peteet*, the plaintiff chose to allege that the other component of Agent Orange, 2,4-D (2,4-dichlorophenoxy acetic acid), caused the decedent's lymphoma. Despite histological evidence that the decedent had Hodgkin's disease and epidemiologic studies that argued against Hodgkin's disease as being associated with exposure to 2,4-D, plaintiff's sole expert was permitted to testify that the decedent died of non-Hodgkin's lymphoma and that just one molecule of 2,4-D was enough to cause the decedent's disease. Moreover, the methods used by the plaintiff's expert were questioned in that he relied upon another person's medical records in forming his opinions. Despite this, the Third Circuit Court of Appeals upheld the verdict for the plaintiff.

*Rubanick v. Witco Chemical Corp., et al.*⁵² is a state court extension of the *Ferebee* concept. In *Rubanick*, the survivors of two deceased chemical company workers brought suit alleging that exposure to chemicals in the work place caused the decedents' cancers. The plaintiffs produced an expert witness who held a Ph.D. in chemistry. This witness testified at a preliminary evidentiary hearing that, based upon the epidemiology of

colon cancer (the alleged cancer in the suit), the personal histories of the deceased, the fact that other workers in the plant developed other types of cancer, and data from animal studies, that decedents' colon cancers were caused by workplace PCB exposure. The defendants produced three experts, a medical doctor coincidentally from the same institution where plaintiffs' expert worked, a toxicologist, and an epidemiologist. These witnesses testified that plaintiffs' expert was unqualified to render opinions as to the specific cause of an individual's cancer, that his theory was "untenable from the scientific standpoint", and that he was not qualified to interpret epidemiologic evidence. At the close of the hearing regarding these matters, the trial court determined that the testimony of plaintiffs' expert was inadmissible and granted a dismissal for the defendant. The intermediate appellate court reversed that decision, however, upon the grounds that the standard for admissibility applied by the trial court was too strict for a toxic tort litigation. They concluded that "... the conventional standard for admissibility of causation ... [requires] a different test of reliability, one focusing on the soundness of the foundation for the novel scientific theory of causation, ... in toxic tort litigation."

The New Jersey Supreme Court agreed. In a decision published on 1 August, 1991, the court determined that toxic tort cases were unique in that much of the science involved with the causes of latent injuries such as cancer is nebulous. The court adopted the position that, in these cases, it is beyond the province of the trial court to determine whether an expert's opinion is one that is generally accepted in the particular field of science. The court further noted that the trial court should hold a hearing to determine whether the methodologies used by the expert in forming his or her opinions are acceptable methodologies and of the type that experts in the field generally use to form opinions. However, the trial court should not judge the opinions themselves; this, said the New Jersey Supreme Court, is the function of the jury.

The court also noted that the expert's credentials should be examined in an evidentiary hearing, but that the jury again should be allowed to weigh an expert's credentials when they are borderline.

Specific criticism was made of those cases wherein judges have excluded expert opinion because the opinion was outside of generally accepted scientific principles. The New Jersey Supreme Court in *Rubanick*, however, did not apply this standard to cases other than toxic tort situations that involve latent injuries and uncertain health effects.

A few months after *Rubanick*, the New Jersey Superior Court was faced with a motion for summary judgment, this time in a pharmaceutical case, *Erickson v. Winthrop Laboratories et al.*⁵³ The plaintiffs in this case were alleging that exposure to 3% hexachlorophene in a skin cleanser caused an infant to suffer mental retardation. After extensive discovery, defendant manufacturers filed for summary judgment on the grounds that

1. the opinions expressed by plaintiffs' experts were not based upon generally accepted methodologies,
2. their opinions were not based on the types of datum normally relied upon by experts in the field, and
3. their opinions were not in line with generally accepted scientific principles, since the conclusions reached by the plaintiffs' experts did not logically flow from the scientific data then available.

Plaintiffs opposed the motion, basing their opposition on *Rubanick*. The plaintiffs claimed that pharmaceutical products should not be differentiated from toxic chemicals as appear in toxic tort cases. Defendants noted that a great difference exists between pharmaceuticals and toxic chemicals, in that pharmaceutical products are specifically tested for their safety on humans prior to use, whereas many toxic tort chemicals are not so tested. Further, they argued that pharmaceuticals are approved by the US Food and Drug Administration for human use, whereas toxic chemicals generally are not. Finally, they argued that the product in question had been on the market for over 40 years without scientific acceptance of any similar reports as was alleged in this lawsuit.

The judge in this case rendered an opinion granting the defendants summary judgment. The

judge agreed with the defendants that pharmaceutical products are not analogous to the types of chemical that appear in toxic tort cases. Further, the alleged injury in this case did not involve any latency. Thus, the court felt that the *Rubanick* standard should not apply. As noted by the court,

The *Rubanick* reasoning leads this court to analyze the question of whether the [sic] this litigation falls within the purview of a toxic tort or a pharmaceutical product case. If it does not fall within the scope of a toxic tort case, the conventional "general acceptance" standard applies.

... In toxic tort cases where PCBs or other known carcinogens are not intentionally administered to humans, study of the carcinogen is difficult because there is scarce information about human exposure to such carcinogens. In contrast, Phisohex is a product that was intentionally administered to humans and has been used by millions of people over the past forty years ... Thus, this court holds that Phisohex is not a toxic tort for purposes of a *Rubanick* analysis. Rather, it is a pharmaceutical drug to which the "general acceptance" standard for evaluating expert testimony applies.

The court in *Erickson* also looked at the methodologies used by plaintiffs' experts in light of *Rubanick* and determined that, even under this standard, the methodologies were flawed and, thus, the testimony of the plaintiffs' experts was inadmissible.

Several other recent cases have excluded testimony of experts because of its unreliability, despite the *Ferebee*, *Paoli*, and *Rubanick* reasoning. These include *Boyles v. American Cyanamid Company*,⁴⁸ *Porter v. Whitehall Laboratories, Inc., et al.*,⁵⁴ and *Kracker v. Spartan Chemical Co., Inc., et al.*⁵⁵

The Supreme Court of the United States decided the case of *Daubert v. Merrill Dow Pharmaceuticals Inc.*⁵⁶ At issue was a Court of Appeals (9th Circuit) ruling affirming summary judgment against plaintiffs who alleged that birth defects were caused by Bendectin.[®] The district court had ruled that the plaintiffs' experts' testimonies were inadmissible because their methodologies were not generally accepted within their scientific disciplines. The Ninth Circuit upheld the decision but proceeded to apply a standard for admissibility which was set

forth in an older case, *Frye v. United States*, 54 App. D.C. 46, 47, 293 F. 1013, 1014 (1923). On appeal to the Supreme Court, the plaintiffs/petitioners contended that the Ninth Circuit erred, in that they applied the *Frye* standard to the Federal Rules of Evidence and that the Federal Rules of Evidence actually replaced the *Frye* standard with their enactment. Defendants/respondents countered by arguing that Federal Rules 403, 702, and 703 allow a judge to consider several factors in determining the admissibility of evidence and that, based upon these rules, summary judgment was justified.

The Supreme Court decision in *Daubert* (No. 92-102, decided June 28, 1993)⁵⁶ took an intermediate position, setting forth a number of parameters that trial courts need to consider in evaluating expert testimony. This decision has had major implications to courtroom use of epidemiologic evidence, since one of the contentions was that an analysis of epidemiologic data performed by plaintiff's experts for purposes of the litigation was not subject to peer review.

Initially, the Supreme Court expressly stated that the *Frye* "general acceptance" test was superseded by the Federal Rules of Evidence and that "general acceptance" of expert opinions is not a precondition to admissibility.⁵⁶ However, the court noted that the Federal Rules, especially Rule 702, do place limits on the admissibility of scientific evidence. Further, the court set forth a number of methods available to a trial judge in assessing the admissibility of expert testimony. Among these are included an evaluation of "... whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology can be applied to the facts in issue ... Whether a theory or technique is scientific knowledge that will assist the trier of fact ... [and] whether it can be (and has been) tested." Proceeding further, the Court opined that publication does not automatically lead to admissibility nor does it necessarily correlate with reliability. However, peer review of a theory or process used by an expert in forming his/her opinion is something the court should consider. The fact that something is too new, novel, or of limited interest for it to be published is

something the trial court should realize but, also, the trial court needs to comprehend the fact that "... submission to the scrutiny of the scientific community is a component of "good science", in part because it increases the likelihood that substantive flaws in methodology will be detected." The Court also pointed out that a trial court, in assessing a particular scientific technique, should consider the known or potential rate of error of the method used. Finally, the Court acknowledged that general acceptance can be "... an important factor in ruling particular evidence admissible, and 'a known technique that has been able to attract minimal support within the community,' may properly be viewed with skepticism."

As mentioned previously, Federal Rule 702 allows an expert to express opinions, or other testimony as to scientific, technical, or other specialized knowledge if it will assist the trier of fact in understanding the data. Rule 403 allows a judge to exclude evidence, even if relevant, if it can mislead the trier of fact. Thus, defendants/respondents in *Daubert* argued that these rules allow a judge to exclude expert testimony if it can mislead a jury and not assist the jury in understanding technical issues. The Supreme Court's decision seems to support this position. However, it is the methods, rather than the opinions, that should be subject to scrutiny, and the Federal Rules of Evidence do require a "gatekeeping role for the judge" in this regard.

The use of experts in courts of law developed because of a need to aid nonexperts who must decide a case in understanding technical issues outside of the common human experience. To lose sight of this function does a disservice to those for whom the use of experts was envisioned. Science requires testing, affirmance, reproducibility by others, verification of data, verification of findings, and a host of other safeguards to protect against misinterpretation and error. Although there are instances of "major breakthroughs" that have occurred outside of the conventional process, such a happenstance is the rarity. However, even after such an event, the "new" finding must withstand scientific scrutiny. The problem with experts testifying in lawsuits to heretofore unknown or

generally unaccepted phenomena is that the safeguards are not in place. Although opposing experts may disagree in front of a jury, technical points may become obscured by other factors (e.g., personalities, testimonial mannerisms, sympathy, etc.), thus allowing a lay jury to be misled.

The decision in *Daubert* enforces the fact that courts should not lose sight of the role of the expert at trial. The Supreme Court made clear that "general acceptance" is not a precondition to the admissibility of scientific evidence. However, it also made clear that to allow experts to mislead is to hinder rather than aid the judicial system.

LITIGATION USE OF SPONTANEOUS REPORTS

The fact that a report is filed about a particular patient noting the association of an adverse event with the use of a defendant's drug raises a possibility of causation, but it is quite insufficient to rise to the level of even a probability. As courts have noted on many occasions, "*post hoc ergo propter hoc*" does not provide sufficient evidence of probability for proof of legal cause. It is legally insufficient to show a possibility that the injury complained of was caused by the product. Justice Griffith, in *Kramer v. Wilkins*, spoke of "a possibility of skin cancer" as not meeting plaintiff's requirement to show causation, when other experts stated it was not a probability. "Possibilities will not sustain a verdict."⁵⁷ The courts have many times held properly that no inference of defectiveness arises from the mere fact that an injury has occurred.⁵⁸ Indeed, it is questionable whether the presence of unverified double and sometimes triple hearsay spontaneous reports, elicited from the present FDA ADE reporting system (see Chapter 10), should suffice even to thwart an early motion for summary judgment by the defendant. These are generally viewed, from scientific and expert pharmacoepidemiologic standards, as insufficient to support the plaintiff's claim. They should be viewed in the same way by the courts.

Yet the FDA makes available, under the Freedom of Information Act,⁵⁹ the ongoing results of its Spontaneous Reporting System. Appropriately

included along with these results are the following caveats:

There are important things to remember when reviewing or analyzing data from the Spontaneous Reporting System:

1. For any given report, there is no certainty that the drug caused the reaction. This is because physicians are encouraged to report suspected reactions. In many instances, the event may have been related to the underlying disease for which the drug was given and not the drug itself.
2. The accumulated case reports can not be inferred to reflect national incidences or estimates of drug risk.
3. Rates and/or frequencies generated from these data must be carefully interpreted as reporting rates and not occurrence rates. True incidence rates cannot be determined from this database.
4. Since reporting is voluntary it depends on many factors. This leads to reporting artifacts which must be kept in mind by anyone reviewing these data. Therefore comparisons of drugs for their relative safety should not be done without a complete understanding of the following and their impact on the comparisons.
 - a. Length of time the drug has been marketed
 - b. Changes in indication for use, or dosage schedules
5. Duplicate reports exist in the file. This must be kept in mind especially when generating and reviewing summary counts. There are two reasons for duplication.
 - a. FDA was unable to determine that a report was previously submitted and computerized.
 - b. A follow-up report was received, identified and entered as a separate case report. This is potentially solvable by examining the accession number on a detailed computer print-out. The last digit of the accession number is "1" for initial reports and a number greater than 1 represents a follow-up report.

This book is mostly concerned with the use of data from formal or otherwise carefully conceived and validated epidemiologic studies derived from biostatistics. To say, however, that such studies may be offered as relevant to the issue of causation is not to say that individual case reports within or without such a study should be, in and of themselves. If such case reports are not acceptable to scientists as such, not only can they not be used

for causation, but they should not be admissible for the purpose of determining sufficiency of warning. Given that pharmacoepidemiology is a mathematical application of the powers and value of medical observation and given also that it can accumulate figures on the occurrence and distribution of adverse drug reactions in established human populations, it has merit for its use by the FDA in determining labeling and in designating products whose right to remain available is suspect. However, the ADE which appears within the FDA's Spontaneous Reporting System is *prima facie* a hearsay or double hearsay document, whose validity singularly is open to considerable doubt. Thus, the use of such as evidence should occasion much need of proof, apart from proof of causation.

TYPICAL LITIGATION USE OF ADE REPORTS

Spontaneous reporting systems are utterly essential to postmarketing surveillance, in particular for hypothesis generation. As outlined in Chapters 10, 11, and 32, however, they are highly flawed for hypothesis testing. Unfortunately, one untoward result of such a system is that the numbers and characteristics it posts for each drug–disease association, because they are recorded, tallied, and distributed as such, create an aura of authenticity that is completely misleading, despite disclaimers.

While not the intention of the FDA Spontaneous Reporting System, materials submitted and accrued in compliance with the statutory and regulatory requirements have become a substantial contributor to drug, device, and vaccine product litigation. A legend appears on the FDA form that used to be used for logging or reporting ADEs (Form 1639) "Submission of a report does not necessarily constitute an admission that the drug caused the adverse reaction." The new MedWatch Form 3500 (see Chapter 10) includes the statement: "Submission of a report does not constitute an admission that medical personnel or the product caused or contributed to the event." No report is intended by the reporter or the FDA to take a definitive position as to causation with respect to the individual experience reflected in the

report. Neither does a tabulation of ADEs referring to a particular product prove positive causality. Medical experts who seek to argue or hypothesize from them in positive terms, no less than attorneys, are open to suspicion, to doubt, and to criticism. However, these materials are available for use by either party, and this may lead to testimony based on these materials.

Regardless of the fact that the FDA and pharmacoepidemiologists generally are dissatisfied with use of the present methods of reporting and recording ADEs to make judgments as to causation, many attorneys persist in introducing them for such purpose in product liability litigation. Thus, the poorest of all compilations, that resulting from the Spontaneous Reporting System of the FDA, often serves the plaintiff's attorney, especially where the judiciary or the defense is equally unsophisticated, much of the means to fend off an initial motion for dismissal and even to make out a *prima facie* case. The availability of these reports is an exciting discovery to the neophyte plaintiff's attorney who has a possible drug product liability lawsuit to consider. By studying the package insert or Physicians' Desk Reference (PDR) monograph for the drug, he can determine whether his client's claim of injury is described. If it is not clear that the claimed injury could possibly result from the use of the target product, a proper inquiry under the Freedom of Information Act can determine for him whether it has been reported. For the neophyte, each of these reports represents a real occurrence of a recorded reaction from the drug. If, as is highly unlikely, it is reported with greater frequency than it is described in the package insert, he has reason to be a little excited. He may be able to show that the labeler was on notice of the danger and did not incorporate it into his labeling in a timely manner. He may be able to show that it happened more frequently and with more serious results than was noted in the labeling. He may be able to show that the manufacturer reported ADEs inaccurately or with great delay, or even that the manufacturer never reported them.

Besides supplying himself with a listing of adverse reactions reported with use of the drug, a plaintiff's attorney can buy a listing with accession numbers for each individual ADE report and

followup report. Reports in the FDA's computerized database of adverse reactions come from a variety of sources, but 80% or better are being forwarded by the regulatee, the product distributor (see Chapter 10). Since these separate reports do not identify the name of the reporting physician and institution, a plaintiff's attorney may want to demand those precise reports from the defendant, in the hope of obtaining such identification. It also is impressive to the defendant to see such diligence. However, there are many problems. Some of these are problems generic to all present adverse reaction reporting. Some of these are problems involving attempts under the regulations to maintain the same confidentiality as does the FDA.

Once the underlying incident has been converted to an actual legal claim, both plaintiff and defendants will look to see whether an ADE has been filed concerning the event. Thereafter, the plaintiff will at some point seek to have an expert examination of the New Drug Application and subsequent file, which will reveal not only all the research and details of FDA approval, but the medical bibliography and ADE information as well. All available information will be examined to prepare the way to further evidentiary exploration, or to prepare for expert testimony on the eventual claim of tortious conduct or product defect.

Thus, a check into the defendant's ADE materials and system may give the plaintiff much capital. When operating in a product liability climate of strict liability, the plaintiff need only convince a jury that the product was rendered defective before administration to the plaintiff. This defect can be inadequacy in product design, in testing and study, in formulation or choice of materials, in provision of instructions for safe and effective use, or in failure to note proper contraindications, warnings, or precautions as to its use. The registry of ADEs is a pathway to the claimed "failure to warn." Given then the ability to show that the product was defective upon administration to the plaintiff, the only other element necessary to the strict liability case is to show a sum of money due to redress the injury caused to the plaintiff by that defect.

Mere association of an ADE with the use of a drug is of itself not a basis for a labeling change.

Similarly, medical events that coincidentally occur along with the use of the drug are not necessarily ADEs from that drug, to be included in the full disclosure labeling for the product. Why exhaust the reader of that labeling with incidental descriptions of unsubstantiated risks, diluting his interest? Meaningful pharmacoepidemiologic findings that support causality assessment in both the medical and legal community are quality observations of associations between the particular product and the particular injury, which are consistently observed by qualified but different observers in a variety of settings. Courts, and the judges whose discretionary power is so vast in the trial courts, must be more sophisticated in determining the credibility of scientific and medical evidence and opinions. They should not allow parties to present unique, non-peer-evaluated reports of studies, and the theories and statistics derived therefrom, to lay juries.

Neither the manufacturer nor the FDA wants every spontaneous report of an ADE, or every bibliographic reference to an ADE, to be incorporated into the package insert. Certainly the prescribing or administering practitioner does not either. Yet, as the manufacturer sensitive to product liability sees the alternatives, he approves the entry of many adverse events, however unvalidated, into his package insert sections for adverse reactions, solely to disarm a later plaintiff who will allege his client's fate would have been different had that notice been available in the package insert before the incident. A court, finding allegation of a similar injury had been made before this one, and not finding warning of it in the package insert although it appears in the manufacturer's adverse reaction files, tends to place much weight on the circumstances in favor of the plaintiff.

IMPACT OF LITIGATION ON PHARMACOEPIDEMOLOGY

Litigation can have a major impact on pharmacoepidemiology, in turn. Frequently involved as codefendants in litigation are others who brought the plaintiff into contact with the product, such as

those who prescribed, administered, dispensed, or otherwise provided the product to the plaintiff. The majority of product liability cases involving prescription drugs also involve claims against treating physicians, hospitals, etc.⁶⁰ Plaintiffs frequently seek to bring the various defendants into conflict with each other, to help make out the case. The perceived threat of being named in litigation, whether or not the threat is real, is a major deterrent to reporting adverse reactions to the FDA's Spontaneous Reporting System. Also, because a practitioner or institution is aware of a potential for litigation at the time they compose the report of the incident, in many instances it is difficult for manufacturers to give as much credence to the resultant spontaneous reports, and often it is fruitless to attempt to follow them up.

Litigation can precipitate or result from in the filing of misleading adverse drug reaction reports. Formerly Form 1639s and currently, MedWatch form 3500s have identified a defendant drug company's product as a culprit, even when it was not used. The reports called into a company by any source, may contain biased, untrue, and prejudicial information. Even attorneys (or their agents) can make complaints that can lead to ADE filing. Thus, courts may properly exclude these reports as prejudicial hearsay, since they may not even rise to the level of putting a company on notice of actual events. It has been the experience of these authors that reports dealing with myelographic media and, most recently, respiratory inhalants fell into this category of erroneous reports and even though the suits were dismissed by the plaintiffs, the reporting forms remain within the FDA files. The problem is compounded by generic drugs. Many times a generic version of a trade named product is dispensed, but the report is made to the trade name manufacturer. At times, this has resulted in lawsuits against the innovator company, even though a generic version was used. Finally, when the generic is identified, the case against the innovator manufacturer may be dismissed, but the ADE form remains. Thus, a skewing of adverse reports biased against a manufacturer's products results. Add to these then a species of report generated from practitioners

threatened with litigation due to their acts and the epidemiologic role of pharmacovigilance is rendered uncertain at best.⁴⁵ This becomes a report aimed to direct the potential plaintiff away from him or his institution, to the “deep pocket” of the manufacturer. Imagine what such a report, and any attempted followup, will then contribute to a pharmacoepidemiologic study!

Lawyers for defendants are frequently appalled to see the misuse of ADE records and reports through the discovery process. We have personally known attorneys to use reports, however unrelated to their own plaintiff’s allegations of injury, as a means to identify physicians and hospital staff for their contact in their own lawsuit. The doctor who has made the report in assured confidence, as spelled out by the regulations, is amazed, embarrassed, and perhaps frightened by a plaintiff’s attorney who recites the doctor’s ADE experience from the “1639” he has “discovered.” This did not result from a breach of confidentiality at FDA, but from a lawyer’s diligent work using the manufacturer’s files during the pretrial discovery process. The call may only be to ask him if he would provide expert testimony for the plaintiff. Alternatively, it may be to enlist the doctor’s antipathy to the product if it exists. However, whatever the purpose, that doctor may think twice before he fills out another ADE!⁶¹

Finally, litigation can affect formal pharmacoepidemiology studies, as well. As the earlier example and those to follow demonstrate, it can generate a major need for pharmacoepidemiology studies. However, it also can create a climate where manufacturers are inappropriately fearful of the results of such studies, and are hesitant to fund them except when they are obligated to do so.

SELECTED EXAMPLES

THE SWINE FLU EXAMPLE

In order to encourage the manufacture and distribution of swine flu vaccine for a federal program of immunization, the government created an extraordinary limitation of future plaintiffs’ rights to sue.⁶² The effectuating law virtually

removed any opportunity for a plaintiff to sue without showing negligence actionable under the Federal Tort Claims Act. Thus, merely showing a product defect, as would be sufficient to support actions in strict liability, would not be enough.

Immediately following the program and also considerably thereafter, claims arose for patient injury in the form of Guillain–Barré syndrome, an acute and serious polyneuritis. Plaintiffs attempted to establish their right to sue the government on the basis of epidemiologic studies of persons receiving the vaccine who later were diagnosed as suffering from the Guillain–Barré syndrome. Like some other syndromes, rather than representing an objectively definable disease entity, an inordinate amount of subjectivity is needed in the diagnosis of this disease. Therefore, a plethora of contradictory epidemiologic analysis was sought to be submitted as evidence by a body of experts, many of whom had little or no formal epidemiologic training.

To defend, and fairly, the government asked the Centers for Disease Control (CDCs) to conduct an epidemiologic analysis of the many records and reports available. Perhaps the most important result of this analysis was the determination that too many variables took over, in time, to consider this analysis as evidence for causation. In light of the circumstances, the CDC proposed, and the federal agency and courts accepted, the premise that only such claims of Guillain–Barré syndrome as arose within 10 weeks or less of the date of injection were credible for litigation.⁶³ All others were deemed improper claimants and only rarely were permitted to litigate to a successful result.⁶⁴

DALKON SHIELD AND THE IUD CALAMITY

In the Dalkon Shield debacle for A.H. Robins Co. Inc., epidemiological testimony was used early on to indicate the level of awareness that might be expected from the product supplier. Thus, although the product was withdrawn from the market in 1974, thousands of suits were filed, posited on Robins’ failure to warn and to acknowledge a purported flawed design, given the multitude of reports that accrued associating it with injuries to users.⁶⁵

DES, GOES ON AND ON

The diethylstilbestrol (DES) litigation is another of the important series of cases, involving billions of dollars, that was created by pharmacoepidemiologic analyses, using both retrospective and prospective studies. To lawyers, it resulted in an important breakthrough in addressing unidentifiable defendants, sparsely identifiable patients, and carcinogenic results, overcoming statutory limitations as to the time for filing suits, and proving defects in products cleared by the FDA.⁶⁶ New theories, as well as expanded old theories, were prepared and accepted by the courts, overwhelmed with the generic nature of DES.

Since all DES was satisfactorily implicated by plaintiff's attorneys, no matter the identity of the actual manufacturer of the pill the mother had ingested, the problem was simplified for the pharmacoepidemiologist. It seems to these authors and others that the effect of the newer amendments to the Federal Food, Drug, and Cosmetic Act to promote the supply of generic products which are deemed bioequivalent will be to create "DES" opportunities for plaintiffs claiming injury from other generic groupings of potent drugs. Thus, the important aspects of all reporting and recording systems for ADEs and postmarketing surveillance will represent an opportunity for the pharmacoepidemiologist to increase the denominators in his or her analysis. It also indicates that cooperation among complainants and, separately, cooperation among defendants, must increase.

The DES litigation also began creating a perhaps inordinate focus by the legal profession on reproductive toxicity. It carries with it a mystical mixture of anger and fear, compounded by frustration about incomplete knowledge about what to fear and what to be angry about. Conveyed, however, to the media, and ultimately to a jury, it is the stuff of which substantial judgments for classes of patients can be fashioned. Toss into the hopper ingredients such as the incalculable insults that chemicals can carry for the individual and the generic complexity of the factors that make up the male and female reproductive system. Add the experience—part documented, part intuitive—that such insults

might be generationally disclosed, or even encoded on genes for later disclosure. Stir in hypotheses from the various medical disciplines, and it forms a strong and interesting incentive for attorneys, as well as scientists, physicians, and public health administrators. We see all of these in play, as well as the combined efforts of federal and state agencies to seek exploration, modification, and regulation in these areas.⁶⁷

What epidemiologic evidence will support the new "third generation" DES cases? "In the DES field," said a leading plaintiff's attorney, "there is a strong possibility that the drug is now reaching its third level of liability ... progeny of injured progeny ... it may be added that there is strong indications that ... patients to whom DES was administered, are sometimes heart victims."⁶⁸ Since in many states cases were permitted to be filed beyond the normal date of expiration by special "toxic tort" legislation, the year revival statutes such as the one signed by Governor Cuomo in New York, attorneys are anxiously bringing suit within the new extended period. While most are relying upon the wealth of materials prepared by plaintiff's attorneys in years past, by dint of a legal doctrine called *collateral estoppel*, both defense of such cases and even extension to the fourth generation lawsuit seems like a clarion call for pharmacoepidemiologic expertise. *Collateral estoppel* is a principle by which a plaintiff can profit from precedential judgments on causation and other important factors, providing these have been definitively established in earlier decisions. A plaintiff is not expected to prove again that which has been proven in a prior case. For example, in New York DES plaintiffs use *Bichler v. Eli Lilly and Co.*⁶⁹ to foreclose attempts by defendants to claim that the plaintiffs failed to provide adequate proof of drug defect or manufacturer fault. Much judicial discretion exists in this regard.

Also, as one contemplates third and even fourth generation cases, consider all the settlements and releases that have been consummated in the prior cases involving the same defendant, probably the same plaintiff's attorney, and representing the same family, even if not the same plaintiff. Releases may or may not, depending on their

exact language, cover claims for injuries discovered after their execution. Yet, as the Indiana Appellate Court has stated it well “Releasees do not make settlement and take releases merely to pay the releaser the first installment ... leaving the matter open for the releaser to come back for more ... On the contrary, a settlement is made and a general release taken for the purpose of foreclosing further claims.”⁷⁰

THE BENDECTIN[®] CONTROVERSY

The controversial use of pharmacoepidemiologic evidence was recently perhaps best reflected in the series of lawsuits that were based on teratogenic manifestations in infants born to mothers who had ingested Bendectin[®] to allay morning sickness while pregnant. Although the manufacturer and many pharmacoepidemiologists believed the drug not responsible, the manufacturer voluntarily withdrew it from the marketplace, a drug that from profit had turned to a potential bankruptcy force.

Teratogenicity is an inflammatory charge and one that may subtly and forcefully affect jury judgment. Reports of birth defects picked up in the FDA’s Spontaneous Reporting System are used by plaintiff’s counsel to keep the case viable so it reaches a jury. Because of the frightening nature of teratogenicity, the plaintiff’s success with the jury can then be anticipated. The most that the defendant can aim for sometimes is reversal of a sympathetic jury verdict. A good example was in *Richardson v. Richardson-Merrell, Inc.*, where the jury found appropriate a judgment of 1.16 million dollars.¹⁴ When the defendant moved for a judgment n.o.v. (i.e., for the judge to cast aside the jury verdict and rule in his favor) or the granting of a new trial, the federal district judge held that no reasonable jury could find on the basis of the scientific evidence that Bendectin[®] “more likely than not” caused the infant plaintiff’s birth defects or that Bendectin[®] was teratogenic (see also previous discussion).

Thalidomide, the catastrophic teratogen that helped to enact the New Drug Amendments of 1962, had been sought to be marketed in the US by then William S. Merrell Labs, whose successor

corporation was defending the Bendectin[®] action. Merrell was defendant in a number of thalidomide suits involving deformed offspring. As might be expected, plaintiff used all but the name thalidomide in the Bendectin[®] case to invoke the past specter of liability for teratogenicity that hovered over the manufacturer. In fact, as this case was presented, “the parties and witnesses in recounting those implications, were required to employ the euphemism MRL-41 for Thalidomide” in an attempt by the court to prevent plaintiffs from invidiously equating the two. However, “MRL” itself was easily recognizable to the jury as Merrell.

Failing to prove an actual product effect that can be termed causative by probability assessment, plaintiffs will seek to elicit a “second best” theory of idiosyncrasy. In the Richardson case, the court refused the suggestion, also offered by the plaintiff to the jury, that even if the vast majority of infants exposed to Bendectin[®] were normal, that Bendectin[®] could be viewed as a “selective” teratogen and that Etheleen (the mother) or Carita (the infant) had a peculiar susceptibility to it (idiosyncrasy), or the composition of the Bendectin[®] distributed in her geographic area must have differed from that distributed elsewhere.

Although plaintiffs in the Bendectin[®] cases have relied upon animal toxicology, *in vitro* tests, and interpretations of Merrell’s animal experiments during the period of investigation, their main argument was that the molecular structure of the antihistamine in Bendectin[®] was similar in some respects to that of antihistamines known to be teratogenic in animals. For example, plaintiff’s witness described the effects of Bendectin[®] components in solution in *in vitro* experiments on frog nerve fibers and mesenchyme cells of mouse limb buds, postulating that other effects might occur in the human intrauterine environment, when Bendectin[®] was infused via the mother, which could inhibit the development of fetal organs. The defendants rebutted by “discipline to discipline” retorts” ... supplemented by extensive clinical experience and testimony.” Experts for the defendant asserted that “structural chemical resemblance does not import similar chemical activity; and that *in vitro* studies of animal tissues have never been scientifically validated as predictors of

physiological effects upon human beings ...” In his comments concerning the plaintiff’s chief principal witness on causation, Judge Jackson indicated the uncertainty of the court in accepting “possibilities” in place of “probabilities.” That expert had no hesitancy in declaring that

Based upon his knowledge of Bendectin[®]’s chemical structure, the *in vitro* studies, the animal teratology studies conducted by Merrell (and others), and the “human data” he had reviewed, i.e., epidemiological studies which he found defective, inconclusive, or both, Dr. Done stated that, in his opinion, to a “reasonable degree of medical certainty,” Bendectin[®] was not only capable of causing birth defects in humans, but that it had, in fact, caused those limb defects with which Carita Richardson had been born.

Commenting directly, the court noted

Had the epidemiological studies reported a statistically significant association between Bendectin[®] and limb reduction defects, Dr Done would have found the causal connection to have been established to a “scientific certainty” as well. As it was, they were unnecessary to his opinion for “medical certainty” purposes. The “drug experience reports” were likewise unnecessary to his opinion, although he would have relied upon them if permitted to do so.

It is interesting to note that, at trial, the same judge took a strong position as to admission and/or reliance upon Drug Experience Reports that plaintiffs had relied upon, among other items, to fight early dismissal. Whether in frustration or irascibility, the judge did not hesitate to imply criticism of plaintiff’s medical expert. And so, in this court, albeit not in all courts,

Plaintiffs were not permitted to introduce into evidence, nor were their experts allowed to rely upon, so-called “drug experience reports” or “adverse reaction reports,” i.e., anecdotal case reports of birth defects observed in the offspring of mothers whose histories included Bendectin[®] usage. The Court made a disputed preliminary finding that such reports are neither exceptions to the hearsay rule nor data reasonably relied upon by experts in the field of making determinations of causality.

At any rate, the court noted that, although the plaintiff’s expert(s) might disagree (citing their various disagreements), the FDA, having convened its advisory committee, evaluated all the evidence, including its own files of adverse drug reaction reports. The panel was composed of “five obstetricians, a biostatistician, a neonatologist, and a “consumer representative”, and was assisted by consultant epidemiologists from the National Institutes of Health and the National Cancer Institute.” It was asked by the FDA to answer three questions:

whether there is, in fact, an increased incidence of birth defects among Bendectin[®]-exposed infants; whether further animal and epidemiological studies were needed; and whether Bendectin[®] should be withdrawn from the market. The answers according to the testimony of Dr David F. Archer, the chairman of the panel, were all in the negative.

The position taken by the judge in his opinion⁷¹ is almost a direct analogy to the Cardozan opinion as to proximate cause more than half century prior, and puts epidemiological analysis on a somewhat different plane than regular evidence, when causation is at issue.

It is necessary to revive debate upon the wisdom of that rule of law which, for purposes of awarding damages in personal injury cases, allows unschooled triers-of-fact, whether judges or juries, to resolve those complex and refractory causal issues ... at the frontier of current medical and epidemiological inquiry.

As the judge opined as to Bendectin[®], the impassioned advancement of pharmacoepidemiologic analyses of case reports as some form of evidence of causation is perhaps based on the “*post hoc ergo propter hoc*” logic, which perverts the use of *res ipsa loquitur*. It argues that the organism is injured and the organism experienced the suspect substance; therefore the suspect substance must have caused the injury. In general, this is the same argument brought to bear in “Agent Orange” product liability litigation.⁷² There again, a federal judge (J. Weinstein) decried reliance on pharmacoepidemiologic conjecture to suit a plaintiff’s theory of causation.

Thus, the Bendectin[®] line of cases demonstrate the two types of plaintiff's attempt to offer proof of causation when a pharmaceutical ingredient is alleged to be the cause of injury, whether that allegation concerns teratogenicity, mutagenicity, or toxicity generally. One, as we have seen in *Oxendine* and in *Richardson* above, is based on laboratory studies, whether in animals or isolated animal tissues, whether *in vivo* or *in vitro*. When pharmacoepidemiologic presentations were somewhat inhibited in *Richardson*, the plaintiff's expert placed considerable reliance on the animal studies and chemical relationships. In "*Agent Orange*," the attempt provoked the opposite response. There Chief Judge Weinstein had noted that studies dealing with animal exposure to Agent Orange were conceded by even the plaintiff's expert to be unpersuasive. As for other offerings that are expected by a party to be probative on the issue of causation, the results of animal testing must first pass a threshold test of reliability and acceptability by recognized authorities in the area of inquiry. In "*Agent Orange*" they did not.

SPERMICIDES AND BIRTH DEFECTS

In contrast, the case *Wells v. Ortho Pharmaceutical Corp.* was the opposite situation to that which occurred in Bendectin[®], where the judge overturned the verdict due to the state of medical science. In *Wells*, the manufacturer's product was implicated as a cause of a limb reduction defect on the basis of a single flawed study, subsequently disproven, cloaked epidemiologically through a rather singular medical opinion. The case had been heard before a Georgia federal judge, who was not swayed by the defendant's arguments of "junk science." He formally ruled that the law need not be bound by medical science. It was impossible for the defendants to appeal, since the US Supreme Court did not accept the case for review.⁷³

THE WOODERSON CASE

Finally, the "Wooderson" case became a cause in which the entire major pharmaceutical industry recognized a stake, and ultimately tendered an amicus brief to the US Supreme Court, but to little

avail. Pharmacoepidemiologists can draw some lessons from this case. No matter how well trained and reputed one may be, a jury can choose to believe one's believed inferiors. Early on, British pharmacoepidemiology data pointed to a need to add a warning about hemolytic uremic syndrome to a certain dose level of hormonal contraceptive pills. FDA and American sources thought the analyses were unreliable and the FDA discouraged the pharmaceutical companies who were marketing such oral contraceptives from incorporating the additional warning. By the time the studies were duplicated in the US, and the package insert changed to incorporate the warning the British had earlier adopted, the plaintiff had developed the exact injury.

Lawyers drew from this case, with its multi-million dollar compensatory and punitive damage awards, the conclusion that, when epidemiologic evidence points to a danger, it may be better to adopt a warning before its more sophisticated validation.⁷⁴

Somewhat similarly, when reports of tetracycline staining infant's teeth were noted, the FDA's approach to epidemiologic analysis left them unconvinced that this was a genuine concern for warning. The tetracycline manufacturers may have enjoyed the FDA sponsored period of "nonwarning" at the time, but it was no favor. Ultimately, many lawsuits were filed and the issue became one of "state of the art" in labeling the product.²⁰

PHISOHEX[®]

Recently completed, the *Stahl* litigation involved the claim that the use of pHisoHex[®] (containing 3% hexachlorophene) as a bacteriostatic skin cleanser to help control impetigo in the diaper area of a two week old, full term infant, resulted in permanent seizure disorder and mental retardation.⁷⁵ The use of pharmacoepidemiologic evidence was very pertinent in the defense of the case. Controversy arose regarding an incident in France, commonly referred to as Talc Morhange. Using epidemiological followup, the defense demonstrated the lack of evidence of permanent seizure disorders and mental retardation in the population that survived that incident. Further, numerous

studies conducted around the world provided evidence against the plaintiff's proposition. (It should be noted that the plaintiff had the burden to prove cause and effect, yet the defendant had to try to prove a negative in order to win.) The defendants put on their own experts, who were able to demonstrate that, despite the widespread use of pHisoHex[®] on infants in millions of American children over the period of 1964 through 1972, there was no observable increase in the incidence of seizure disorders. Thus, the lack of such a finding was demonstrative of the lack of credible evidence in support of plaintiff's contention. After a 3-month trial, a defense verdict was rendered by the jury in less than two hours.

CONCLUSIONS

Thus, the legal profession must join with professional colleagues in medicine and contributing sciences to assure that, in the best interests of the public, the reporting and recording procedures created to assure the safety and efficacy of therapeutic and diagnostic modalities are maintained and improved. They must also seek to discourage misuse or distortion of the information gleaned from these procedures, just as they must encourage the development and inclusion of "state of the art" warnings and instructions that accompany these products to the public.

As advice for educators, we have previously recommended recognition that the needs of the pharmacoepidemiologists' three major constituencies—public health agencies, academia, and industry—probably cannot be met by a unitary teaching and training experience. As pointed out in the address of one of the authors to the Third International Conference on Pharmacoepidemiology in 1987, we would suggest that each university program in pharmacoepidemiology arrange for a course at the university or another affiliated law school. This course should explore the legal burden born by governmental agencies in implementing their statutory mandate. It should also provide a brief excursion into medicolegal and pharmacolegal principles associated with litigation involving both health professionals and suppliers

of health care products. We would also suggest that, within the university curriculum in pharmacoepidemiology, an "internship" be offered, that would place candidates for certification as epidemiologists in an industrial habitat of their choice. In response, numerous eligible companies would register with the university, indicating their needs, their resources, and their capacity for affording the "intern" with the additional training to which he/she aspires.

Finally, as advice for manufacturers, the incorporation of "pharmacoepidemiology interns," plus ongoing state of the art formal pharmacoepidemiology studies, would be in keeping with the invaluable pharmacoepidemiology prophylaxis that we envision to be part of appropriate planning by pharmaceutical manufacturers for the defense of subsequent litigation. As adopted by many now, it consists of either "in-house" or independent epidemiologists, functioning as outside consultants, conducting periodic or ongoing reviews of pharmacoepidemiologic data concerning the product line. This, plus any ongoing formal pharmacoepidemiology studies, can then be presented at trial as indicative of company diligence and intentions, as well as providing rationale for the currency of the labeling of the product in use at the time of the incident that generated the complaint.

REFERENCES

1. 611 F. Supp 1223 (1985) *in re* Agent Orange Litigation. E.D.N.Y.
2. Braun v. Roux 312 S.W. 2nd 758 (Missouri, 1958).
3. Section 505 of the Federal Food, Drug, and Cosmetic Act 21 USC 321 *et seq.*
4. Section 505(b) of Federal Food, Drug, and Cosmetic Act 21 USC 321 *et seq.*
5. Section 505(c) of Federal Food, Drug, and Cosmetic Act 21 USC 321 *et seq.*
6. Section 505(e) of Federal Food, Drug, and Cosmetic Act 21 USC 321 *et seq.*
7. Section 505(j) of Federal Food, Drug, and Cosmetic Act 21 USC 321 *et seq.*
8. 21 CFR 310 *et seq.*
9. Willig SH, Tuckerman MM, Hitchings WS. *Current Good Manufacturing Practices for the Pharmaceutical Industry* 2nd edn. New York: Dekker, 1982.
10. US v. Parkinson, 240 F.2nd 918, 9th Circuit, (1956).

11. 21 USC 301, 22 USC 332, 21 USC 333, 21 USC 334.
12. Makuc v. Honda 835 F.2d 392 (1987).
13. Calvanese v. Babcock 412 N.E. 2nd 895, Mass (1980).
14. Richardson v. Richardson–Merrell Inc. 649 F. Supp 799 (1986).
15. Oxendine v. Merrell Dow (District of Columbia) 506 A.2d 1100 (1986).
16. Learned intermediary rule applies to medical devices. Phelps v. Sherwood Medical Industries and Argyle Associates 836 F.2d 296. Kirk v. Michael Reese Hospital 117 111. 2nd 507 (1987).
17. Reyes v. Wyeth 498 F.2d 1264 (1974).
18. Davis v. Wyeth 399 F.2d 121 (1968).
19. Lindsay v. Ortho 637 F.2d 87 (1980).
20. Feldman v. Lederle Lab 97 N.J., 429, 429 A.2d 374 (1984).
21. Willig SW. the comment k concept, *Mercer Law J* 1978; Spring; 545–81. See also Dalko v. Upjohn 555 F.2d 245 (1977).
Simms v. S.W. Texas Methodist Hospital 535 SW 2nd 192 (1976).
22. Eutsler v. Orthopedic Equipment Co. 276 F.2d 455, VA (1960).
23. Leveck v. Consolidated Rail Corp, 498, NE 2nd 529 Ind. (1986).
24. F.E.L.A. (Federal Employees Liability Act) 45 USC 51 et seq.
25. Summar v. I.H. Belt R. Co. 498 NE 2nd 828 Ind (1986).
26. Toole v. Richardson-Merrell, Inc. 251 Cal. App. 2nd 689, 60 Cal Rptr. 398 (1967).
27. Love v. Wolf, 226 Cal. App. 2nd 378, 38 Cal Rptr. 183 (1964).
28. Mulder v. Parke Davis & Co. 288 Minn. 332, 181 NW 2nd (1970).
Annot. 45 ALR 3rd 920 (1972).
28. Ogilvie v. Int. Playtex, Inc. 821 F. 2nd 1438 (Kansas), (1987).
Carter v. Hoffman La Roche Inc. 492 P.2nd 1000, New Mexico, 1971.
29. Stanton v. Astra, 718 F.2d 553 US Court of Appeals, Third Circuit (1983).
30. McCall v. Sisson 336 NE 2nd 660, IND (1975).
31. *In re* “Agent Orange” Product Liability Litigation 603 F. Supp 239 (1985).
32. Hutchinson A. On Flaws in Epidemiologic Approach. *Drug Inf Assoc J* 1986; 20: 439.
33. Justice McFadden, Supreme Court of Idaho, in “Henderson” 518 P.2nd 873, Idaho (1974).
34. Noel DW, Phillips JJ. *Products Liability* 2nd edn. St. Paul, MN: West, 1976; 526.
35. Carter v. Hoffman-La Roche Inc. 492 P.2nd 1000, New Mexico, 1971.
36. Justice Cardozo cited in MacPherson v. Buick, 217 N.Y. 382, 111 NE 1050 (1916).
37. Palsgraf v. L.I.R.R. 248 NY 33 9.
38. Dore M. A proposed standard for evaluating the use of epidemiological evidence. 1985; *Howard University Laws J* 28: 677.
39. Willig SH. *Address to 3rd International Congress on Pharmacoepidemiology*, University of Minnesota, 1987.
40. 611 F. Supp 223, 1241 (NY) 1985; affirmed 818 F.2d 145 (NY) 1987.
41. Johnston v. The United States of America, *et al.*, 597 F. Supp. 374 (D. Kansas, 1984).
42. Allen v. The United States of America, 588 F. Supp. 247 (C.D. Utah, 1984).
43. *In re* “Agent Orange” Product Liability Litigation, 611 F. Supp. 1223 (E.D.N.Y., 1984), affirmed., 818 F.2d 187 (2nd Cir., 1987), cert. denied 487 US 1234 (1988).
44. Viterbo *et ux.* v. The Dow Chemical Co., 646 F. Supp. 1420 (E.D. Tex., 1986), affirmed., 826 F.2d 420 (5th Cir., 1987).
45. Thomas v. Hoffman-LaRoche, Inc. 731 F. Supp. 224 (N.D.Miss., 1989), affirmed., 949 F.2d 806 (5th Cir., 1992).
46. Novak v. The United States of America, 865 F.2d (6th Cir., 1989).
47. Sterling v. Velsicol Chemical Corp., 855 F.2d 1188 (6th Cir., 1988).
48. Boyles v. American Cyanamid Co., 87 CV 3357 (E.D.N.Y., 1992).
49. Ferebee v. Chevron Chemical Co., 736 F.2d 1529, cert. denied, 469 US 1062.
50. *In re* Paoli Railroad Yard PCB Litigation, 1988 WL 128399 (E.D. Pa., 1988), 916 F.2d 829 (3rd Cir., 1990), cert. denied, 113 L.Ed.2d 649 (1991), E.D. Pa. No. 86–2229, Oct. 1992.
51. Peteet v. The Dow Chemical Co., 868 F.2d 1428 (5th Cir., 1989), affirmed., 493 US 935 (1989).
52. Rubanick v. Witco Chemical Corp., *et al.*, 125 N.J. 421 (1991).
53. Erickson v. Winthrop Laboratories, *et al.*, N.J. Super. Ct., No. L-071373–84, Aug. 7, 1992.
54. Porter v. Whitehall Laboratories, Inc., *et al.*, 1992 WL 92764 (S.D. Ind.).
55. Kracker v. Spartan Chemical Co., Inc., *et al.*, N.Y.App. Div., 2nd Dept., No. 91–04737, 1992.
56. Daubert v. Merrill Dow Pharmaceuticals Inc., 727 F. Supp. 570 (S.D.Cal., 1989), affirmed., 951 F.2d 1128 (9th Cir., 1991), cert. granted 113 S.Ct. 320 (1992), decided, June 28, 1993, US Sup. Ct. No. 92–102.
57. Kramer v. Wilkins, 186 So. 625, Miss (1939).
58. Phillips v. US Waco Corp; 516 NE 2nd 670, Illinois (1987).
59. Freedom of Information Act Amendments of 1974. See 39 ed. Reg. 446 02, December 24, 1974 for FDA implementation.
60. Willig SW: “Partners in litigation”. *Mercer Law Journal* 1986; 37: 755–88. See also Goellner v. Julian

- Butler, M.D., University of Minnesota. 836 F.2d 426.
61. 21 CFR 314.81(h) has other intentions.
 62. Swine Flu Act, 42 USC 247b(j)–(1). Enacted 1976. For background, see Baumes TE Jr. Reflections on the swine flu experience; drug manufacturer's liability for failure to warn. 1977; *St Louis U Law J* 21: 44.
 63. Francis Szczezanizk Jr. v. USA, Civil Action No. 80-990-M.A., Mar 2, 1983.
 64. Baker v. US 660 F. Supp. 204 (1987).
 65. *In re* A.H. Robins Co. Inc. 406 F. Supp 540. See also 419 F. Supp 710 (1976), 438 F. Supp 942 (1977); 453 F. Supp. 108 (1978); 505 F. Supp 211 (1981); 570 F. Supp. 1480 (1983).
 66. Sindell v. Abbott Labs. 26 Cal 3rd 588, 163 Cal. Rptr. 132; 607 P.2nd 924.
 67. Fed Reg. 43; 7174–7179.
 68. Julien AS. *New York Law J.* 1987; 198 (126): 1.
 69. Bicular v. Eli Lilly and Co., 55 NY 2nd 571, 450 NYS 2nd 776 (1982).
 70. Hybarger v. A.S. Ins. Co., 498 N.E. 2nd 1019 (1986).
 71. 649 F. Supp 803. See also 14 *supra*.
 72. 611 F. Supp 1223 E.D. NY (1985).
 73. Wells v. Ortho Pharmaceutical Corp., 615 F. Supp 262, 788 F. 2nd 741, 107 Supreme Court 437, cert. denied (1987).
 74. Wooderson v. Ortho 235 Kansas 387, 681 F.2nd 1038; cert. denied 105 S.(t 365) (1984).
 75. Stahl v. Rhee, Ban and Sterling Drug, Inc. (643 N.Y.S. 2d 148), Suffolk County, NY, Verdict (1999). Also, *New York Law J* 1999; 1.

Part III

SYSTEMS AVAILABLE FOR
PHARMACOEPIDEMIOLOGY STUDIES

10

Spontaneous Reporting in the United States

DIANNE L. KENNEDY, STEPHEN A. GOLDMAN AND RALPH B. LILLIE
Food and Drug Administration, Rockville, MD, USA

INTRODUCTION

The United States Food and Drug Administration (FDA) has the regulatory responsibility for ensuring the safety of all marketed pharmaceuticals (i.e., drugs and biologics). While the US has one of the most rigorous pre-approval processes in the world, clinical trials cannot uncover every safety problem, and they are not expected to do so. Premarketing clinical trials are effective tools primarily designed for assessing efficacy and risk–benefit ratio. However, due to the limited size and controlled nature of these studies, only the most common adverse events will be observed and subsequently listed in the product's official labeling at the time of approval. The need for postmarketing surveillance is a direct result of these limitations. A degree of uncertainty always exists about both the benefits and risks from pharmaceuticals. The tradeoff for accepting these uncertainties is continued vigilance by the FDA and industry to collect and assess data during the postmarketing life of a product.

The FDA monitors the quality of marketed pharmaceuticals through product testing and surveillance. In addition, the agency develops policies, guidances, and standards for product labeling, current good manufacturing practices,

clinical and good laboratory practices and industry practices to demonstrate the safety and effectiveness of pharmaceuticals.

As shown in Figure 10.1, risks from medical products generally fall into four categories.^{1,2} Most injuries and deaths associated with the use of medical products result from their *known side-effects*. Some side-effects are unavoidable even when the product is used appropriately (e.g., nausea with chemotherapy), but others can be prevented or minimized by careful product choice and use. It is estimated that more than half the side-effects of pharmaceuticals are avoidable.³ Other sources of preventable adverse events are *medication errors*, which may occur when the product is administered incorrectly or the wrong drug or dose is administered. Injury from *product defects* is unusual in the United States, because of the great attention paid to product quality control and quality assurance during manufacturing. The final category of potential risk involves the *remaining uncertainties* about a product. These include unexpected adverse events (AEs), long term effects, and unstudied uses and/or unstudied populations.

The FDA collects and maintains reports of AEs as an important component of its postmarketing surveillance program. These are sent to the FDA voluntarily from a variety of sources and as part of

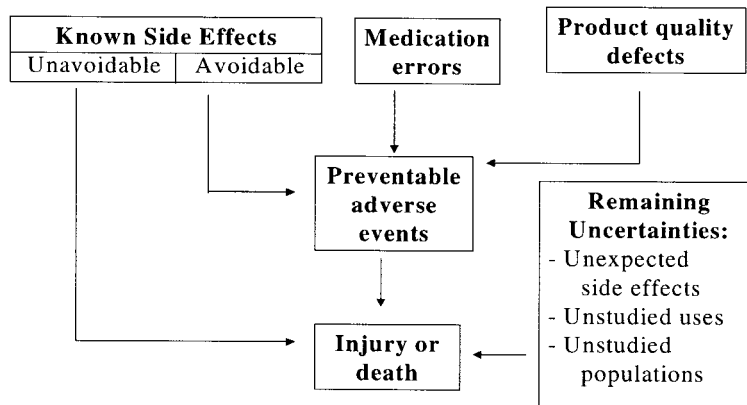


Figure 10.1. Sources of risk from medical products.

a regulatory reporting requirement of pharmaceutical manufacturers. The term *spontaneous reporting* itself is somewhat misleading. Spontaneous reporting is spontaneous only in the sense that health professionals, consumers, or other individuals who initially decide to contact a manufacturer about a suspected adverse event or report an AE directly to the FDA do so at their own discretion. Thus, spontaneous generally refers to AEs *spontaneously observed* during the usual practice of medicine. However, the FDA's Adverse Event Reporting System (AERS) also contains AE reports that are a result of observations made during more formal clinical studies, referred to as "study reports," and reports from the medical and scientific literature.

Spontaneous AE reporting focuses on unexpected AEs, and may supply some information regarding unstudied uses and/or unstudied populations, but its effectiveness in detecting long term effects is less certain. The principal use of this type of passive reporting is to detect serious, unknown, rare events. Different methodologies for collecting AEs are being explored regarding the other remaining uncertainties.

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 are discussed, as are future plans.

DESCRIPTION

HISTORY OF US PHARMACEUTICAL SAFETY REGULATION

The development of the US national postmarketing system, unfortunately, has been based on a series of medical disasters. The first US drug law came into existence during the 1848 Mexican War, after drugs imported for the US Army were found to have been adulterated.⁴ This law, although poorly enforced, outlawed the importation of such adulterated drugs.

The Progressive Era legislation that addressed drug safety was the Federal Food and Drugs Act of 1906, which defined and forbade the misbranding or adulteration of foods and drugs. However, the safety of drugs after consumption was not addressed until the 1930s, when there was a drive to enact legislation to update the 1906 food and drugs statute. It took a convincing disaster, however, to prompt the US Congress to include drug safety provisions in the new law.⁵

During October and November of 1937, 107 people in the US died after exposure to Elixir Sulfanilamide. Neither the safety nor efficacy of the sulfanilamide was to blame. It was the solvent used in the preparation of the so-called elixir, diethylene glycol, which was the poison.⁶ 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, because there was no alcohol in the “elixir” as implied by the name.

In the aftermath of this disaster, new drug safety regulations were placed in the bill and the Food, Drug, and Cosmetic Act of 1938 was passed by the US Congress. The legislation required the submission of a New Drug Application (NDA), in which proof of drug safety was required for approval. Drugs on the market before 1938 were “grandfathered” with respect to drug safety testing, i.e., they were not required to submit, and thus do not have, an approved NDA.

In the early 1950s, reports of aplastic anemia associated with chloramphenicol use demonstrated the necessity for monitoring adverse events following the approval and marketing of new drugs. Chloromycetin[®] had passed drug safety testing for approval in 1949; however, the small number of patients exposed to chloramphenicol before approval was not adequate to observe this rare adverse event.⁷

In response to the need for a system to detect rare adverse events that do not appear until after drug approval, the American Medical Association (AMA) established a Committee on Blood Dyscrasias, which began collecting case reports of drug induced blood related illness in 1952. This group was later expanded to monitor all AEs.

During the 1950s, there was a rapid expansion of the pharmaceutical industry and an increase in the number of new products. Concerns began to be raised over drug efficacy claims. Once again, a disaster had to occur before legislation was passed. The well known tragedy of 10 000 birth defects (worldwide) as a result of *in utero* exposure to thalidomide⁸ led the US Congress to address the need for more comprehensive drug regulation. 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, they also required pharmaceutical manufacturers to report AEs to the FDA for any of their products having an NDA, which constituted the vast majority of prescription products introduced since 1938.

The FDA began computer based storage of its AE reports in the mid-1960s, with data now retrievable only from late 1969 because of incompatibilities with early data files. The first computerized data storage system was called the Spontaneous Reporting System (SRS). The SRS was replaced in November 1997 with the new Adverse Event Reporting System (AERS), an internationally compatible system designed as a pharmacovigilance tool.

REGULATORY REPORTING REQUIREMENTS

In the US, AE reporting by individual healthcare providers or consumers is voluntary. However, manufacturers, packers, and distributors of FDA approved pharmaceuticals (drugs and biological 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.⁹

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.

Current Requirements

The main objective of the FDA postmarketing reporting requirement is to provide 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 1998.¹⁰

An *adverse experience* (AE) is any adverse event associated with the use of a drug or biological 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 pharmacological 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, be life-threatening, or require hospitalization may be considered a serious AE when, based upon appropriate medical judgement, 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 10.1 outlines US mandatory reporting regarding pharmaceuticals. By regulation, these companies are required to report *all* adverse events of which they are aware to the FDA and to provide as complete information as possible. Although pharmaceutical reporting is mandated, it still relies heavily on information provided 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,¹¹ the manufacturers of these latter products do not have to prove safety or efficacy, but that same law puts the onus on the FDA to prove that a particular product is unsafe. In addition, manufacturers of these products do not have to report AEs to the FDA. As a result, direct-to-FDA voluntary health professional reporting of serious adverse events possibly associated with dietary supplements is particularly important.

Table 10.1. Mandatory 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 Report" 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 three 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, and must be translated into English if foreign.

Postmarketing studies. No requirement for a 15 day Alert Report on an AE acquired from a postmarketing study *unless* manufacturer concludes a reasonable possibility that the product caused the event.

The specific regulations governing postmarketing AE reporting by pharmaceutical companies are listed in Table 10.2. Accompanying separate guidances for drugs and biologics were made available in 1992¹² and 1993¹³ respectively. As can be seen, the regulations have each been amended numerous times.

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 Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). These efforts toward international harmonization have a direct impact on how FDA is currently rewriting regulations on adverse event reporting. The Adverse Event Reporting System (AERS), launched November 1997 and evolving, is an internationally compatible system in full accord with the ICH initiatives.

The initiatives that directly affect postmarketing surveillance are the following.

- *M1 IMT (International Medical Terminology)*. AERS uses *MEDDRA (Medical Dictionary for Drug Regulatory Affairs)* as its primary tool to classify and search for medically significant adverse events.
- *M2 ESTR I (Electronic Standards for the Transfer of Regulatory Information)*. AERS will require the use of ESTR I standards for submission of adverse reaction reports in electronic form.
- *E2A (Clinical Safety Data Management)*. AERS implements the harmonized postmarketing surveillance definitions contained in E2A.¹⁴
- *E2B (Data Elements For Transmission of Individual Case Safety Reports)*. AERS has implemented the E2B data format into its database, and will require manufacturers to use the E2B standard for electronic submissions.
- *E2C PSUR (Periodic Safety Update Report)*. Defines a standard format for periodic reports that manufacturers can submit to all ICH member regulatory authorities. Initially, PSURs will be submitted on paper, and will not be processed within AERS.

In line with international activities, the FDA has undertaken a major effort to clarify and revise its regulations regarding postmarketing safety reporting requirements for pharmaceutical products.

The agency published a proposed rule in the *Federal Register* of 27 October, 1994 (59 FR 54046) to amend these requirements, to implement international standards, and to facilitate the reporting of AEs. In the *Federal Register* of 7 October, 1997 (62 FR 52237), the FDA published

Table 10.2. Federal regulations regarding postmarketing AE reporting

<p>21 CFR 310.305 Prescription drugs not subject to premarket approval [3 Jul, 1986 (51 FR 24779), amended 13 Oct, 1987 (52 FR 37936); 29 March, 1990 (55 FR 1578); 28 Apr; 1993 (57 FR 17980); 25 June, 1997 (62 FR 34167); 7 Oct, 1997 (62 FR 52249)]</p> <p>21 CFR 314.80 Human drugs with approved new drug applications (NDAs) [22 Feb, 1985 (50 FR 7493) and 11 Apr, 1985 (50 FR 14212), amended 23 May, 1985 (50 FR 21238); 3 July, 1986 (51 FR 24481); 13 Oct, 1987 (52 FR 37936); 29 March, 1990 (55 FR 11580); 28 April, 1992 (57 FR 17983); 25 June, 1997 (62 FR 34166, 34168); 7 Oct, 1997 (62 FR 52251); 26 Mar, 1998 (63 FR 14611)]</p> <p>21 CFR 314.98 Human drugs with approved abbreviated new drug applications (ANDAs) [28 Apr, 1992 (57 FR 17983), amended 8 Sept, 1993 (58 FR 47352)]</p> <p>21 CFR 600.80 Biological products with approved product license applications (PLAs) [27 Oct, 1994 (59 FR 54042), amended 25 June, 1997 (62 FR 34168); 7 Oct, 1997 (62 FR 52252); 26 Mar, 1998 (63 FR 14612)]</p>	<p>● <i>M1 IMT (International Medical Terminology)</i>. AERS uses <i>MEDDRA (Medical Dictionary for Drug Regulatory Affairs)</i> as its primary tool to classify and search for medically significant adverse events.</p> <p>● <i>M2 ESTR I (Electronic Standards for the Transfer of Regulatory Information)</i>. AERS will require the use of ESTR I standards for submission of adverse reaction reports in electronic form.</p> <p>● <i>E2A (Clinical Safety Data Management)</i>. AERS implements the harmonized postmarketing surveillance definitions contained in E2A.¹⁴</p> <p>● <i>E2B (Data Elements For Transmission of Individual Case Safety Reports)</i>. AERS has implemented the E2B data format into its database, and will require manufacturers to use the E2B standard for electronic submissions.</p> <p>● <i>E2C PSUR (Periodic Safety Update Report)</i>. Defines a standard format for periodic reports that manufacturers can submit to all ICH member regulatory authorities. Initially, PSURs will be submitted on paper, and will not be processed within AERS.</p>
---	---

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 timeframe, and harmonizes reporting pre- and postmarketing and internationally/domestically. To help the pharmaceutical manufacturers understand the new requirements, on 27 August, 1998 the FDA published an interim guidance for industry—"Postmarketing adverse experience reporting for human drugs and licensed biological products: clarification of what to report."

In the *Federal Register* of 5 November, 1998 (63 FR 59746), the agency published an advanced notice of proposed rule making 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 ICH in the M1, M2, and E2B initiatives. Many pharmaceutical companies are currently participating in an AERS electronic submission pilot project to jointly identify and resolve any technical or process issues related to the transfer of data via this means.

In addition, following a presidential directive to eliminate or modify regulations that are outdated or in need of reform, the FDA published a final rule in the *Federal Register* of 25 June, 1997 (62 FR 34166) that revokes the postmarketing safety reporting requirement to submit 15 day increased frequency reports for serious and expected AEs. This action was based on FDA's determination that expedited increased frequency reports have not contributed to the timely identification of safety problems requiring regulatory action and are no longer necessary for FDA surveillance of postmarketing adverse experiences.

Proposed Modifications

At the current time, the FDA is working on further modifications to the postmarketing safety reporting

requirements. An Adverse Drug Reaction Reporting Proposed Rule will be published that focuses on report quality, standardizes terminology to "Adverse Drug Reaction," encourages active query, followup, and determination of seriousness, defines the minimum data set for safety reports, provides specific followup timeframes, and encourages a clinically intelligent process. The proposed rule will also implement the ICH E2C: International PSUR, which contains marketing status, core labeling, changes in safety status since last report, exposure data, clinical explanation of cases, data line listings and tables, status of postmarketing surveillance safety studies, overall critical analyses, and assessments. With regard to the 27 October, 1994 proposed amendments to the postmarketing periodic AE reporting requirements, these amendments will be repropounded in this proposed rule based on a guidance on this topic developed by ICH ("Guideline on clinical safety data management: periodic safety update reports for marketed drugs"; 19 May, 1997 (62 FR 27470)).

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. This rule will propose the requirement for ADR reporting for OTC monograph products and amend certain ADR reporting requirements for OTC drugs with approved NDAs.

Along with the changing regulations, the Agency will publish an overall "Guidance for industry: adverse experience reporting for human drug and licensed biological products including vaccines" that will supersede the March 1992 "Guideline for postmarketing reporting for adverse drug experiences," the October 1993 "Guideline for adverse experience reporting for licensed biological products," and the August 1997 "Postmarketing adverse experience reporting for human drug and licensed biological products: clarification of what to report."

DATA COLLECTION: THE MEDWATCH PROGRAM

An effective national postmarketing surveillance system is dependent on voluntary reporting of

adverse events and product 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.¹⁵

Some organizations do, however, help facilitate reporting of AEs to the FDA. Adverse event monitoring by hospitals is linked to Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) standards. In order to be accredited, JCAHO requires each hospital 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.¹⁶ The American Society of Health-System Pharmacists has also issued guidelines on ADR monitoring and reporting.¹⁷

Given the vital importance of postmarketing surveillance, MedWatch, the FDA Medical Products Reporting Program, was established in 1993.¹⁸ While FDA’s long-standing postmarketing surveillance program predates MedWatch, this educational/promotional initiative was designed to emphasize the responsibility of healthcare providers to identify and report adverse events and problems related to the use of medical products.¹⁹

The program has four general goals, with the first 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 an active continuing education program that provides free continuing education credit. The continuing education articles are accessible via the MedWatch homepage (<http://www.fda.gov/medwatch>).

The second goal of MedWatch is to clarify what should (and should not) be reported. Health professionals are asked to limit reporting to serious AEs. This is important both in improving the quality of individual reports and 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 for a health professional to notify the FDA and/or the manufacturer.

For those manufacturers participating in the FDA’s “MedWatch to Manufacturer” (MMP) program, copies of serious reports submitted directly to the FDA are sent expeditiously to the manufacturer. More information on the MMP program can be found at <http://www.fda.gov/medwatch/report/mmp.htm>.

The third goal is to make it as easy as possible for a health professional to report directly to the agency. Only one form is necessary for reporting adverse events and product problems with any human use medical product 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 (Figures 10.2 and 10.3). The postage-paid FDA Form 3500 is used for voluntary reporting, while the FDA Form 3500A is used for mandatory reporting, as described earlier. These forms are available from the FDA via a toll-free number (1-800-FDA-1088) or can be printed off the MedWatch homepage or downloaded as a software package.

In addition to mailing in a voluntary form, reporters also have the options of reporting via an interactive form on the MedWatch website, faxing their reports (1-800-FDA-0178), or calling 1-800-FDA-1088 to give 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. VAERS is a joint FDA/Center for



For VOLUNTARY reporting by health professionals of adverse events and product problems

Form Approved: OMB No. 0910-0291 Expires: 7/31/99 See OMB statement on renewal

FDA Use Only

Triage unit sequence #

Page ___ of ___

A. Patient information

1. Patient identifier In confidence 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/yr) 4. Date of this report (mo/day/yr)

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) - -

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 & address 4. Operator of device [] health professional [] lay user/patient [] other: _____

5. Expiration date (mo/day/yr)

6. model # 7. If implanted, give date (mo/day/yr)

catalog # 8. If explanted, give date (mo/day/yr)

serial # other #

lot # 9. Device available for evaluation? (Do not send to FDA) [] yes [] no [] returned to manufacturer on (mo/day/yr)

10. Concomitant medical products and therapy dates (exclude treatment of event)

E. Reporter (see confidentiality section on back)

1. Name & address phone #

2. Health professional? [] yes [] no 3. Occupation 4. Also reported to [] manufacturer [] user facility [] distributor

5. If you do NOT want your identity disclosed to the manufacturer, place an " X " in this box. []

PLEASE TYPE OR USE BLACK INK



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

ADVICE ABOUT VOLUNTARY REPORTING

Report experiences with:

- medications (drugs or biologics)
- medical devices (including in-vitro diagnostics)
- special nutritional products (dietary supplements, medical foods, infant formulas)
- other products regulated by FDA

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

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

- suspected contamination
- questionable stability
- defective components
- poor packaging or labeling
- therapeutic failures

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)

Important numbers:

- 1-800-FDA-0178 to FAX report
- 1-800-FDA-7737 to report by modem
- 1-800-FDA-1088 to report by phone or for more information
- 1-800-822-7967 for a VAERS form for vaccines

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.

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

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:

DHHS Reports Clearance Office
Paperwork Reduction Project (0910-0291)
Hubert H. Humphrey Building, Room 531-H
200 Independence Avenue, S.W.
Washington, DC 20201

"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."

Please DO NOT RETURN this form to this address.

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

FDA Form 3500-back

Please Use Address Provided Below – Just Fold In Thirds, Tape and Mail

Department of Health and 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 Medical Products Reporting Program
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20852-9787



Figure 10.2. MedWatch Voluntary Reporting Form (FDA Form 3500).



For use by user-facilities, distributors and manufacturers for MANDATORY reporting

Form Approved: OMB No. 0910-0291 Expires: 7/31/99 See OMB statement on reverse

Mfr report #
UF/Dist report #
FDA Use Only

Page ___ of ___

A. Patient information

1. Patient identifier In confidence
2. Age at time of event: or Date of birth:
3. Sex: female, male
4. Weight: lbs, kgs

B. Adverse event or product problem

1. Adverse event and/or Product problem
2. Outcomes attributed to adverse event: death, life-threatening, hospitalization, disability, congenital anomaly, required intervention to prevent permanent impairment/damage, other

3. Date of event (mo/day/yr)
4. Date of this report (mo/day/yr)

5. Describe event or problem

Large text area for describing the event or problem.

6. Relevant tests/laboratory data, including dates

Text area for relevant tests and laboratory data.

7. Other relevant history, including preexisting medical conditions

Text area for other relevant history and preexisting conditions.

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) #1, #2

4. Diagnosis for use (indication) #1, #2
5. Event abated after use stopped or dose reduced #1, #2

6. Lot # (if known) #1, #2
7. Exp. date (if known) #1, #2
8. Event reappeared after reintroduction #1, #2

9. NDC # - for product problems only (if known) #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 & address
4. Operator of device: health professional, lay user/patient, other

5. Expiration date (mo/day/yr)

6. model #, catalog #, serial #, lot #, other #

7. If implanted, give date (mo/day/yr)

8. If explanted, give date (mo/day/yr)

9. Device available for evaluation? (Do not send to FDA) yes, no, returned to manufacturer on

10. Concomitant medical products and therapy dates (exclude treatment of event)

Text area for concomitant medical products and therapy dates.

E. Initial reporter

1. Name & address, phone #

2. Health professional? yes, no
3. Occupation
4. Initial reporter also sent report to FDA yes, no, unk

PLEASE TYPE OR USE BLACK INK



FDA Form 3500A

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

Medication and Device Experience Report

(continued)

Submission of a report does not constitute an admission that medical personnel, user facility, 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 ____

FDA Use Only

F. For use by user facility/distributor—devices only

1. Check one <input type="checkbox"/> user facility <input type="checkbox"/> distributor		2. UF/Dist report number	
3. User facility or distributor name/address			
4. Contact person		5. Phone Number	
6. Date user facility or distributor became aware of event (mo/day/yr)		7. Type of report <input type="checkbox"/> initial <input type="checkbox"/> follow-up # _____	
8. Date of this report (mo/day/yr)		9. Approximate age of device	
10. Event problem codes (refer to coding manual)			
patient code		device code	
11. Report sent to FDA? <input type="checkbox"/> yes _____ (mo/day/yr) <input type="checkbox"/> no		12. Location where event occurred <input type="checkbox"/> hospital <input type="checkbox"/> outpatient diagnostic facility <input type="checkbox"/> home <input type="checkbox"/> ambulatory surgical facility <input type="checkbox"/> nursing home <input type="checkbox"/> outpatient treatment facility <input type="checkbox"/> other: _____ specify	
13. Report sent to manufacturer? <input type="checkbox"/> yes _____ (mo/day/yr) <input type="checkbox"/> no		14. Manufacturer name/address	

G. All manufacturers

1. Contact office – name/address (& mfring site for devices)		2. Phone number	
4. Date received by manufacturer (mo/day/yr)		3. Report source (check all that apply) <input type="checkbox"/> foreign <input type="checkbox"/> study <input type="checkbox"/> literature <input type="checkbox"/> consumer <input type="checkbox"/> health professional <input type="checkbox"/> user facility <input type="checkbox"/> company representative <input type="checkbox"/> distributor <input type="checkbox"/> other: _____	
6. If IND, protocol #		5. (A)NDA # _____ IND # _____ PLA # _____ pre-1938 <input type="checkbox"/> yes OTC product <input type="checkbox"/> yes	
7. Type of report (check all that apply) <input type="checkbox"/> 5-day <input type="checkbox"/> 15-day <input type="checkbox"/> 10-day <input type="checkbox"/> periodic <input type="checkbox"/> Initial <input type="checkbox"/> follow-up # _____		8. Adverse event term(s)	
9. Mfr. report number			

H. Device manufacturers only

1. Type of reportable event <input type="checkbox"/> death <input type="checkbox"/> serious injury <input type="checkbox"/> malfunction (see guidelines) <input type="checkbox"/> other: _____		2. If follow-up, what type? <input type="checkbox"/> correction <input type="checkbox"/> additional information <input type="checkbox"/> response to FDA request <input type="checkbox"/> device evaluation	
3. Device evaluated by mfr? <input type="checkbox"/> not returned to mfr. <input type="checkbox"/> yes <input type="checkbox"/> evaluation summary attached <input type="checkbox"/> no (attach page to explain why not) or provide code: _____		4. Device manufacture date (mo/yr)	
5. Labeled for single use? <input type="checkbox"/> yes <input type="checkbox"/> no		6. Evaluation codes (refer to coding manual)	
method		results	
conclusions		7. If remedial action initiated, check type <input type="checkbox"/> recall <input type="checkbox"/> notification <input type="checkbox"/> repair <input type="checkbox"/> inspection <input type="checkbox"/> replace <input type="checkbox"/> patient monitoring <input type="checkbox"/> relabeling <input type="checkbox"/> modification/adjustment <input type="checkbox"/> other: _____	
8. Usage of device <input type="checkbox"/> initial use of device <input type="checkbox"/> reuse <input type="checkbox"/> unknown		9. If action reported to FDA under 21 USC 360i(f), list correction/removal reporting number:	

10. <input type="checkbox"/> Additional manufacturer narrative		and/or		11. <input type="checkbox"/> Corrected data	
--	--	--------	--	---	--

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:

DHHS Reports Clearance Office
Paperwork Reduction Project (0910-0291)
Hubert H. Humphrey Building, Room 531-H
200 Independence Avenue, S.W.
Washington, D.C. 20201

"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."

Please DO NOT RETURN this form to this address.

Figure 10.3. MedWatch Mandatory Reporting Form (FDA Form 3500A).

Disease Control and Prevention program for mandatory reporting by physicians of vaccine related adverse events.²⁰

The fourth goal of MedWatch is to provide feedback to health professionals about new safety issues involving medical products. As new information (e.g., "Dear Health Professional Letters," Public Health Advisories, Safety Alerts, etc.) becomes available, it is posted on the MedWatch Homepage and immediate email notification goes out to members of MedWatch's listservs. To join the general MedWatch listserv, send a subscription message to fdalists@archie.fda.gov. In the message body enter: `subscribe medwatch YourEmailAddress@YourDomain` (replace "YourEmailAddress@YourDomain" with your own email address). MedWatch also has a network of more than 140 health professional and industry organizations that have allied themselves with the FDA as MedWatch Partners. These partners are used to further disseminate safety notifications to their membership, thus acting as multiplier groups to expand notification.

These MedWatch partners encourage all health-care providers (physicians, pharmacists, nurses, dentists, and others) to look upon adverse event reporting as part of their professional responsibility. The American Medical Association and American Dental Association advocate (respectively) physician and dentist participation in adverse event reporting systems as an obligation.^{21,22} Further, since 1994, *The Journal of the American Medical Association* instructs its authors that adverse drug or device reactions should be reported to the appropriate government agency, in addition to submitting such information for publication.²³ The International Committee of Medical Journal Editors recently revised the "Uniform requirements for manuscripts submitted to biomedical journals" to also encourage timely reporting of urgent public health hazards.²⁴

The FDA's interest in informing health professionals about new safety discoveries 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.

The FDA acknowledges that health professionals have concerns regarding their confidentiali-

ty 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. This was further strengthened on 3 July, 1995, when a regulation went into effect extending this protection against disclosure by preempting state discovery laws regarding voluntary reports held by pharmaceutical, biological, and medical device manufacturers.²⁵ To facilitate obtaining followup information, health professionals reporting directly to the FDA are asked to indicate whether they prefer that their identity not be disclosed through the MMP program to the manufacturer of the product involved in the case being reported. When such a preference is indicated, that information will not be shared.

DATA MANIPULATION: THE ADVERSE EVENT MONITORING SYSTEM (AERS)

A new tool used by FDA is the computerized spontaneous adverse event reporting system or AERS, which replaces older technology. This client server, Oracle based system contains all AE reports on pharmaceuticals submitted to the Agency either directly or via the manufacturer.

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

- Improve the operational efficiency, effectiveness, and quality control of the process for handling AEs.
- Improve the accessibility of AE information to all users.
- 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 AE reports to the FDA central document room, where they are sorted and forwarded to the Office of Postmarketing Drug Risk Assessment (OPDRA)

in FDA's Center for Drug Evaluation and Research (CDER). Manufacturer reports are submitted in duplicate, with one copy forwarded to the reviewing division that is responsible for the marketing status of the drug being reported upon. Reports submitted by individuals are mailed, faxed, sent via the internet, or phoned into MedWatch, which in turn forwards the appropriate reports to OPDRA.

When received by OPDRA, these incoming 3500 and 3500A reports are assigned a single permanent report number, imaged and stored in an Excalibur 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 reports receive priority handling.

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 AERS is currently done through manual coding, AERS is designed for electronic submission of ICH standardized, MEDDRA precoded 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 is in place. The pilot program is designed to develop and test the necessary processes, procedures, and technical architecture to implement electronic submissions. The process of transitioning to electronic reporting is a complex one that will take place over the next few years.

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.²⁶

A variety of technology assisted features in AERS augment AE review by OPDRA's safety evaluators. Some of these features are in place and some will be implemented when the system is fully developed. Safety evaluators will 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 designated medical events known to often 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.
- *Graphic monitoring indicators.* Allow visualization of graphical displays to enable signal generation. Eventually this will include statistical tools for data-mining events within the entire database.
- *Periodic (canned) reports.* Enable 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.

When fully deployed, the AERS system will constitute a world-class system for maximizing the ability of the agency to identify and assess signals of importance in the spontaneous reporting system.

FDA EVALUATION OF REPORTS OF ADVERSE EVENTS

The uncontrolled nature of spontaneously reported data places great importance on the complex, intensive process of individual evaluation of submitted AE reports. This is perhaps most accurately characterized as a method, applied on a case-by-case basis, that is based on experience, knowledge of the medical product, and awareness of the limitations of the data.

Given the large number of AE reports that the FDA receives, it is obviously not possible to review all reports individually. The reporting forms seek information on dechallenge and rechallenge (positive, negative, not applicable), and this information is included in the computerized database. However, the causal relationship between the drug exposure and the adverse event reported to the FDA is not individually evaluated for most of the AE reports. Elements such as the temporal relationship between the event and the exposure and possible confounding factors such as concomitant drugs and underlying disease are assessed only for reports of specific AEs that are being closely monitored.

All reports sent directly to the FDA (i.e., nonmanufacturer reports) and all 15 day reports (manufacturer reports involving serious unlabeled events) are reviewed on a case-by-case basis by a safety evaluator, usually a pharmacist or a physician, because they are the most likely to involve events with potential significant public health impact. If the safety evaluator determines that the reported AE is new (i.e., unlabeled) and potentially significant, (s)he will then query the AERS database for similar reports to develop a case series. Each case report is evaluated for the adequacy of the information on the report, the temporal association of the drug and the event, possible confounding factors such as patient disease or concomitant drug therapy, and dechallenge–rechallenge information. If necessary, followup information is requested from the reporter or manufacturer. Further steps such as literature searches, use of drug utilization databases, or conducting further epidemiologic investigations may be taken. If the case series meets certain specific criteria, the issue is brought to the attention of others within OPDRA. An important new event is referred to as a *MAR* (monitored adverse reaction). These *MARs* are given further evaluation and are brought to the attention of the medical review divisions at the FDA that are responsible for regulatory actions involving the drug's marketing status. The reviewing division may request manufacturer sponsored postmarketing studies to further evaluate the issue.

After a decision has been made regarding a safety issue, the FDA can initiate various actions,

including labeling, name, or packaging change(s); a "Dear Health Professional" letter; restricting the use of the drug; or working with a manufacturer regarding possible withdrawal of a medical product from the US market.

To notify health professionals of important new safety information discovered after marketing, the FDA often requests that the manufacturer send a "Dear Health 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.

There were 34 drug or biologic letters/safety notifications posted in 1998, with another 19 added through the end of June 1999. In 1998, safety related labeling changes were approved by FDA for an average of 28 drug products each month (range = 18–41). (Note: "Dear Health Professional" letters and other safety notifications, and summaries of safety related labeling changes approved each month, can be found on the MedWatch homepage (www.fda.gov/medwatch/safety.htm.) Table 10.3 lists some examples of recent "Dear Health Professional" letters.

The FDA can restrict or limit the use of product if the adverse reaction associated with a drug product has severe consequences. For example, fetal abnormalities associated with the use of Accutane[®], a drug to treat acne, resulted in a contraindication for use in women of child-bearing age. If potential safety problems are an issue when a product is first approved, FDA may ask the company to conduct postmarketing studies (phase IV) or to continue phase III trials beyond their planned completion date. For example, results showed that patients taking flosequinan (Manoplax[®]) for congestive heart failure had initial improvement. However, the patients in the study continued to be followed after the product was approved in December 1992. It was ultimately shown that treated patients had an increased risk of death, as well as an increased hospitalization rate.²⁷ This information led to the withdrawal of the drug after just 7 months on the market.

Previously unsuspected safety issues discovered after marketing can also lead to the removal of a

Table 10.3. Recent “Dear Health Professional” letters

<p><i>June 1999: Cylert[R] (pemoline).</i> Updated recommendations for liver function monitoring and a patient information/consent form.</p>
<p><i>June 1999: Rezulin[R] (troglitazone).</i> Based on further evidence of serious and sometimes fatal liver injury, significant new changes in the labeling and recommended uses.</p>
<p><i>June 1999: Trovan[R] (trovafloxacin/alatrofloxacin).</i> Because of associated serious liver injury, use should be reserved for use only in the treatment of patients who meet all of certain specified treatment criteria.</p>
<p><i>June 1999: Propulsid[R] (cisapride).</i> Two new contraindications (known family history of congenital long QT syndrome and clinically significant bradycardia) and a new drug interaction (coadministration of grapefruit juice increases the bioavailability of cisapride and concomitant use should be avoided).</p>
<p><i>May 1999: Celebrex[R] (celecoxib).</i> In patients concurrently taking warfarin, infrequent reports of increases in prothrombin time, sometimes associated with bleeding events, predominantly in the elderly.</p>
<p><i>May 1999: Enbrel[R] (etanercept).</i> Reports indicate that certain patients have developed serious infections, including sepsis, and that several of these patients have died from their infections.</p>
<p><i>March 1999: Xeloda[R] (capecitabine).</i> Reports of altered coagulation parameters and/or bleeding in cancer patients who were taking coumarin derivatives concomitantly.</p>
<p><i>January 1999: Cerebyx[R] (fosphenytoin).</i> Massive overdoses associated with serious adverse events, including death, have resulted from health care workers’ mistaken interpretation of the current vial label (both 2 ml and 10 ml vials). Specifically, some health care workers withdrawing fosphenytoin from the vial have misinterpreted the vial label to mean that the amount of phenytoin equivalents per milliliter actually represents the total amount of phenytoin equivalents in the vial.</p>
<p><i>January 1999: Trovan[R] (alatrofloxacin mesylate).</i> Potential incompatibility of alatrofloxacin mesylate injection with two commonly used diluents, 0.9% sodium chloride injection, USP (usually referred to as normal saline solution), and lactated Ringer’s, USP.</p>
<p><i>January 1999: Flovent[R] (fluticasone propionate).</i> Rare cases of patients on inhaled fluticasone propionate presenting with systemic eosinophilic conditions, some with clinical features of vasculitis consistent with Churg–Strauss syndrome.</p>

drug from the market. Fortunately, such product withdrawals are very uncommon; there have been only 13 drugs taken off the US market since 1980; these are listed in Table 10.4.

In addition to the technology used in current AE 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 OPDRA, the European Union, the World Health Organization (WHO), and others, to systematically screen large databases of spontaneous reports.

OPDRA is specifically testing and validating the feasibility of using data-mining techniques employing a Bayesian statistical basis²⁸ and commercial visualization computer tools to enhance

surveillance efficiency. The strategy is to use the 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 followup.

STRENGTHS

LARGE-SCALE AND COST-EFFECTIVE

Two vital advantages of surveillance systems based on spontaneous reports are that they potentially

Table 10.4. Withdrawals from the market, 1980–1998

Brand name (generic name)	Reason for withdrawal	Year of marketing	Year withdrawn
Selacryn[R] (ticrynafen)	liver necrosis	1979	1980
Oraflex[R] (benoxaprofen)	liver necrosis	1982	1982
Zomax[R] (zomepirac)	anaphylaxis	1980	1983
Merital[R] (nomifensine)	hemolytic anemia	1985	1986
Suprol[R] (suprofen)	flank pain syndrome	1986	1987
Enkaid[R] (encainide)	excessive mortality	1987	1991
Omniflox[R] (temafloxacin)	hemolytic anemia	1992	1992
Manoplax[R] (flosequinan)	excessive mortality	1992	1993
Redux[R] (dexfenfluramine)	cardiac valvular disease	1996	1997
Pondimin[R] (fenfluramine)	cardiac valvular disease	1973	1997
Seldane[R] (terfenadine)	drug interactions/fatal cardiac arrhythmias	1985	1998
Posicor[R] (mibefradil)	multiple drug interactions	1997	1998
Duract[R] (bromfenac)	serious hepatotoxic effects	1997	1998

maintain ongoing surveillance of all patients, and are relatively inexpensive.²⁹ In fact, they are probably the most cost-effective way to detect rare, serious 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.³⁰ Toward that goal, the great utility of spontaneous reports lies in *hypothesis generation*,³¹ with need to explore possible explanations for the adverse event in question. By fostering suspicions,³² 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 10.5 lists factors that are helpful in evaluating the strength of association between a drug and a reported adverse event.³³ (See also Chapter 32.)

The stronger the drug–event relationship in each case and the lower the incidence of the AE occurring spontaneously, the fewer case reports are needed to perceive causality.³⁴ 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.³⁵ In fact, it has been suggested that a temporal relationship between medical product and adverse event, coupled with positive dechallenge and rechallenge, can make isolated reports conclusive as to a product–event association.³⁶ Biological plausibility and reasonable strength of association aid in deeming any association as causal.³⁷

However, achieving certain proof of causality through adverse event reporting is unusual. Attaining a prominent degree of suspicion is much more likely, and may be considered a sufficient basis for regulatory decisions.³⁴

Table 10.5. 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 (*dechallenge*) or continued
- Etiologic roles of agents and diseases in regard to adverse event
- Response to readministration (*rechallenge*) of agent
- Laboratory test results
- Previously known toxicity of agent

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 was the most effective source of new ADR reports that led to changes in labeling.³⁶ Ensuring that the information provided in the AE report is as complete and in depth as possible further enhances postmarketing surveillance. Thus, while possessing inherent limitations, postmarketing surveillance based on spontaneous report data is a powerful tool for detecting AE signals of direct clinical impact.

WEAKNESSES

As with clinical trials, there are important limitations to consider when using spontaneously reported AE information. These limitations include difficulties with AE recognition, underreporting, biases, estimation of population exposure, and report quality.

ADVERSE EVENT RECOGNITION

The recognition of AEs (or any other medical product associated adverse event) is quite subjective and imprecise.³⁸ While an attribution between the medical product and the observed event is assumed with all spontaneously reported events, every effort is made to rule out other explanations for the event in question. It is well known that placebos³⁹ and even no treatment⁴⁰ 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 medica-

tion, alcohol, or “recreational” drug use had caused hospitalization.⁴¹

Such considerations emphasize the crucial need for careful, thoughtful review of adverse event reports upon their receipt by FDA or the manufacturer. It is through this process that causality, or at least a high degree of suspicion for a product–AE association, is put to the test. Ultimately, formal pharmacoepidemiology studies are usually needed though, to be certain.

UNDERREPORTING

Another major concern with any spontaneous reporting system is underreporting of adverse events.^{20,31,37,42} 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.³⁷ A similar estimate is that the FDA receives by direct report less than 1% of suspected serious ADRs.⁴² 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

Unlike clinical trial data, which are obtained under strictly controlled conditions, spontaneously reported information is uncontrolled, and therefore subject to the possible influence of a number of biases that can affect reporting. These biases include the length of time a product has been on the market, country, reporting environment, detailing time, and quality of the data.⁴³ A striking illustration of the impact one such factor can have is the finding that the peak of spontaneous ADR reporting for a drug is at the end of the second year of marketing, with a subsequent precipitous decline in reporting,⁴⁴ despite a lack of apparent decline in usage or change in ADR incidence.^{43,44} 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.³⁵

ESTIMATION OF POPULATION EXPOSURE

Compounding these limitations is the lack of denominator data, such as user population and drug exposure patterns,³⁵ that would provide the exact 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,³⁵ 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.⁴⁵

This approach, whose basic methodologies are applicable to medical products in general, can be of great 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.⁴⁵ Cooperative agreements and contracts with outside researchers enable FDA to use such databases in its investigations. Device utilization studies employ the same sources of data, as well as Medicare derived information.

Care must be taken in interpreting results from studies using these databases. That drug prescribing does not necessarily equal drug usage,⁴⁵ 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 AE 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 AE, 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 (as applicable)
- biopsy/autopsy results (as applicable)
- dechallenge/rechallenge information (if available)
- outcome

SUMMARY

In 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 most underrepresentative of adverse events occurring 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, 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 Adverse Event Reporting System (AERS) contains almost 2 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 1985, as can be seen in Figure 10.4. By 1990, the annual number of reports had risen to 83 000, and in 1998 it was over 230 000.

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 of reports are sent by manufacturers (over 90% in 1998).

Historically, most AE reports have originated with health professionals. However, the proportion coming from consumers has been increasing. In 1993, the proportions from health professionals and consumers were 72 and 27% respectively (1% unknown). In 1996, the proportions were 58, 41, and 1%. Over the four year period from 1993 to 1996, reports from consumers increased in absolute numbers as well as proportionally. Over the same time, although the proportion from health professionals decreased, the absolute numbers went up.⁴⁶ The increase in reports from consumers may very well be related to the large numbers of former prescription products now available OTC, as OTC packaging will frequently contain the company's 800 number.

Although the majority of the FDA's adverse drug event reports are channeled through the manufacturer, a substantial number of reports are received directly from health care professionals (over 15 400 reports in 1998). Direct reports have been shown to be more useful than other types of reports for identifying previously unsuspected AEs,⁴⁷ resulting in the emphasis on such direct reporting via the MedWatch program. Direct reports tend to be more detailed, contain more specific information, and are more likely to have a suspected high likelihood of association between the event and the drug in question. Two-thirds of all direct drug reports involve serious AEs. Compared with all US physicians, physicians who report AEs to the FDA are likely to be younger, to practice in a primary care specialty, and to spend more time in teaching and research activities.⁴⁸

Of reports sent directly to FDA, 73% involve drugs, 14% medical devices, 3% biologics, and 2% dietary supplements. Fifty-nine% are from pharmacists, 15% from physicians, 9% from nurses, and 6% from non-health professionals.

SPECIFIC EXAMPLES

Intravenous Immunoglobulin and Aseptic Meningitis Syndrome

In early 1994, FDA learned of a report from the National Institutes of Health (NIH), which described a high rate of aseptic meningitis syndrome

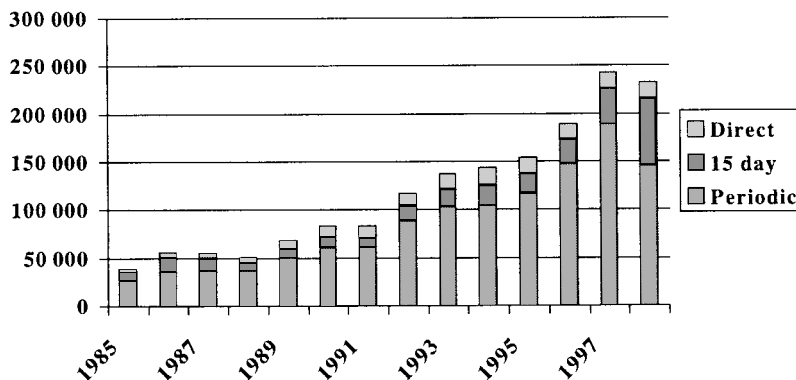


Figure 10.4. ADE reports by year and type 1985–1998 (1998 estimated)

occurring in patients being treated for neuromuscular diseases with high doses of intravenous immunoglobulin (IVIG). The patients had been receiving doses of 2 g/kg of IVIG, which is five to ten times higher than the normally recommended dosage. Six of 54 patients developed severe headache, meningismus, and fever within 24 hours of dosing. Cerebrospinal fluid (CSF) was consistent with aseptic meningitis syndrome in four of the six. Following this lead, 22 cases of IVIG associated aseptic meningitis syndrome that had been reported to the FDA were reviewed. Symptoms included fever and photophobia, and prominent painful headache. Twenty of the cases were associated with positive CSF findings, including leukocytosis (predominantly neutrophilic) and elevated protein.

Unexpectedly, 19 of the reports indicated that normal doses of IVIG had been administered (0.2–0.4 g/kg). The patients had been treated by withdrawal of the medication and administration of analgesics. Of particular note was the characteristic time course of IVIG associated aseptic meningitis syndrome. The illnesses all began between 12 and 24 hours after administration, and recovery ensued within several days following withdrawal of the medication.

As a result of this work, FDA and NIH workers published two articles on IVIG–aseptic meningitis syndrome simultaneously in the same journal.^{49,50} The FDA also directed IVIG manufacturers to modify labeling to include a precaution statement about the occurrence of the syndrome.

Temafloxacin and Hemolytic Anemia

Temafloxacin, a fluoroquinolone antibiotic, was first marketed in January 1992. By early April, FDA had received a few reports of hemolytic anemia occurring in patients treated with this drug. Over the next two months, many additional cases were reported, eventually totaling nearly 100. These provided a clear picture of what was subsequently called the “temafloxacin syndrome.”

The typical patient was a young woman with no underlying medical conditions who was treated for urinary tract infection with temafloxacin. Within 7–10 days of starting treatment, dark colored

urine was often noted, sometimes with accompanying flank pain and chills. There was typically a drop in hemoglobin of 3 grams or greater. Acute renal failure developed in nearly two-thirds, with hemodialysis usually required. Mild hepatobiliary changes were noted in half the patients, and coagulopathy in one-third.

A subset of patients experienced the syndrome after their first dose of temafloxacin. That these patients were more likely to have had prior exposure to a fluoroquinolone antibiotic provided support for an antibody-mediated basis for massive hemolysis.

Based on spontaneously reported cases, the manufacturer, in consultation with FDA, voluntarily withdrew temafloxacin from the market worldwide in June, barely six months after initial marketing. In 1994, FDA staff published a multi-case review article describing the “temafloxacin syndrome.”⁵¹

L-tryptophan Related Eosinophilia–Myalgia Syndrome

In July 1989, a healthy 44-year-old woman in Santa Fe with a history of allergic rhinitis started taking L-tryptophan, an essential amino acid available as a dietary supplement, for insomnia. By early September, she was reporting onset of cough, shortness of breath, and weakness. When first seen by a physician in late September, she presented with a puffy, flushed face, abdominal pain, mucosal ulcers, myalgia, and weakness. Her white blood cell count was 11 900 cells/mm³, with an eosinophil count of 42%. Her condition worsened through October, with her white blood cell rising to 18 200 and eosinophil count to 45%.

Her physician consulted with a rheumatologist, who, while not knowing what was wrong with this patient, did know of a second patient who had been hospitalized in Santa Fe with similar symptoms and eosinophil count. In mid-October, a third patient in New Mexico, who had an eosinophil count of 9000 and had also been taking L-tryptophan, was discovered. While one patient was unusual and two was suspicious, three made it a cluster of a very uncommon disease.

All three original patients were middle-aged women. Although the severity differed, all had the common features of myalgia, weakness, oral ulcers, abdominal pain, shortness of breath, and skin rash. While the doses of L-tryptophan they had used were similar, the duration of use before onset of illness varied from a few weeks to 2 years. Common laboratory features included striking leukocytosis, eosinophilia, elevated aldolase (with a normal creatine kinase), and abnormal liver function tests.

An article about the condition appeared in the 7 November *Albuquerque Journal News*. On 11 November, FDA issued a Public Advisory against the use of L-tryptophan, followed four days later by the establishment by the Centers for Disease Control and Prevention (CDC) of a system of national state based surveillance for the newly named eosinophilia–myalgia syndrome (EMS).⁵²

On 17 November, FDA requested a nationwide recall of all OTC dietary supplements in capsule or tablet form providing 100 mg or more of L-tryptophan in a daily dose. On 23 March, 1990, because of the identification of one case of EMS associated with a dietary supplement containing less than 100 mg, and continued efforts by some firms to circumvent the recall, the agency requested an expansion of the recall to all marketed products containing added manufactured L-tryptophan. Excepted were those that were permitted to contain added L-tryptophan under existing food additive regulations. Additionally, on 22 March, the agency had imposed an import alert to detain all foreign shipments of manufactured L-tryptophan.

Because virtually all manufactured L-tryptophan is imported into the US, the practical effect of the recall and import alert was to effectively eliminate the availability of L-tryptophan containing dietary supplements. Eventually, more than 1500 cases of EMS, including 38 deaths, were reported to the CDC, although the true incidence of the disorder is thought to be much higher.

The recognition of a cluster of cases was the key to the detecting of EMS. Interactions among various specialists, including a family physician, hematologist, rheumatologist, clinical immunologist, and epidemiologists, were crucial to this process.

Of equal importance is ongoing basic and clinical research to explain the etiology and pathogenesis of this disorder. Although it is widely believed that contaminants or impurities in the L-tryptophan are responsible for EMS, continuing research indicates a role for “pure” tryptophan itself,^{53–55} as well as for certain host factors in the etiology of the disorder.^{56,57} These findings support suggestions that the L-tryptophan associated EMS was caused by several factors and is not necessarily related to a contaminant in a single source of L-tryptophan.

FDA concerns about the safety of L-tryptophan containing products and the possibility of potential new cases of L-tryptophan related EMS are underscored by recent information indicating the availability of L-tryptophan by American sources. Both EMS’s clinical seriousness, and uncertainties surrounding its etiology, indicate the need for health professionals to remain vigilant regarding adverse events possibly associated with the use of L-tryptophan containing dietary supplements, and to report such events to MedWatch.

THE FUTURE

Adverse event reporting in the United States has come a long way since its inception in the early 1950s. Yet, it should be remembered that spontaneous reporting, although invaluable, is only one tool used in managing medical product risk. The FDA is committed to evaluating and improving its role in the current risk management system. The May 1999 report to the FDA Commissioner “Managing the risks from medical product use: creating a risk management framework”² found that the postmarketing surveillance program currently in place performs well for the goal it was designed to achieve, namely the rapid detection of unexpected serious AEs that occur postmarketing. However, the report also stated that FDA’s programs are not designed to evaluate the rate, or impact, of known AEs. 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.

In recognition of the increasing importance of postmarketing surveillance and risk assessment in the regulatory setting, a variety of initiatives are under way within FDA to expand these activities. In 1998, the Office of Postmarketing Drug Risk Assessment (OPDRA) was established within CDER to explore and confirm safety signals from spontaneous reporting; to plan, direct, and collaborate in the conduct of epidemiological activities; and to explore other drug surveillance strategies.

Many of the basic problems with spontaneous reporting are still with us today, such as under-reporting and poor quality of reporting. However, FDA, academia, and industry are making efforts to address these problems. These efforts can be summarized in the following areas: increasing the quality of incoming reports; 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.

Studies to explore the barriers to reporting are being proposed by FDA with the goal of increasing the quality of reports and identifying areas in which effective reporting can be stimulated. Specific areas for consideration include the following.

- Identification of causes and factors contributing to product related injuries.
- Factors involved in identification and reporting of AEs.
- Identification of more effective risk communication methods to transmit information to health care professionals and consumers.
- Identification of methods to focus more attention on medical products in the immediate postmarketing period.
- Existing barriers to reporting in areas such as managed care.

In summary, the spontaneous reporting of AEs provides an important cornerstone for pharmacovigilance in the United States. Regulators and manufacturers of medical products worldwide are moving forward 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 to improve our overall ability to manage risk from pharmaceuticals.

REFERENCES

1. Food and Drug Administration, Center for Drug Evaluation and Research. *CDER 1998 Report to the Nation—Improving Public Health through Human Drugs*. Rockville, MD: CDER, 1999.
2. Report to the FDA Commissioner from the Task Force on Risk Management. *Managing the risks from medical product use*. Rockville, MD: Food and Drug Administration, 1999.
3. Bates DW, Leape LL, Petrycki S. Incidence and preventability of adverse drug events in hospitalized adults. *J Gen Intern Med* 1993; **8**: 289–94.
4. Young JH. *Pure food: securing the federal food and drugs act of 1906*. Princeton, NJ: Princeton University Press, 1989.
5. Jackson CO. *Food and drug legislation in the New Deal*. Princeton, NJ: Princeton University Press, 1970.
6. Erslev AJ, Wintrobe MM. Detection and prevention of drug-induced blood dyscrasias. *J Am Med Assoc* 1962; **181**: 114–9.
7. Maeder T. *Adverse reactions*. New York: Morrow, 1994.
8. Schardein JL, ed. *Chemically induced birth defects*. New York: Dekker, 1985; 215–59.
9. Adverse experience reporting requirements for licensed biological products; final rule. *Fed Regist* 1994; **59**: 54 034–44.
10. Expedited safety reporting requirements for human drug and biological products; final rule. *Fed Regist* 1997; **62**: 52 237–53.
11. Dietary Supplement Health and Education Act (DSHEA) of 1994.

12. Food and Drug Administration, Center for Drug Evaluation and Research. *Guideline for postmarketing reporting of adverse drug experiences*. Rockville, MD: CDER, 1992.
13. Food and Drug Administration, Center for Biologics Evaluation and Research. *Guideline for adverse experience reporting for licensed biological products*. Rockville, MD: CBER, 1993.
14. International Conference on Harmonization; Guideline on Clinical Safety Data Management; Definitions and Standards for Expedited Reporting; Availability—Notice. *Fed Reg* 1995; **60**: 11 283–7.
15. Safe Medical Devices Act of 1990, Public Law 101–629, *Stat* 1990; **104**: 4511.
16. Joint Commission on Accreditation of Healthcare Organizations. *Comprehensive Accreditation Manual for Hospitals—updated as of April 1999*. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations, 1999.
17. ASHP guidelines on adverse drug reaction monitoring and reporting. *Am J Health Syst Pharm* 1995; **52**: 417–9.
18. Kessler DA. Introducing MedWatch: a new approach to reporting medication and device adverse effects and product problems. *J Am Med Assoc* 1993; **269**: 2765–8.
19. Goldman SA, Kennedy DL. MedWatch FDA's Medical Products Reporting Program. *Postgrad Med* 1998; **103**: 13, 16.
20. Chen RT, Rastogi SC, Mullen JR, Hayes SW, Cochi SL, Donlon JA, *et al*. The Vaccine Adverse Event Reporting System (VAERS). *Vaccine* 1994; **12**: 542–50.
21. American Medical Association. Reporting adverse drug and medical device events: report of the AMA's Council on Ethical and Judicial Affairs. *Food Drug Law* 1994; **49**: 359–66.
22. Advisory Opinion No. 1 to Section 4-A, Devices and Therapeutic Methods, of the American Dental Association Principles of Ethics and Code of Professional Conduct.
23. *The Journal of the American Medical Association*. JAMA instructions for authors. *J Am Med Assoc* 1998; **280**: 19–26.
24. Glass RM. Reporting of public health hazards or major advances—revision of uniform requirements. *J Am Med Assoc* 1998; **280**: 2035.
25. Protecting the identities of reporters of adverse events and patients; preemption of disclosure rules. *Fed Regist* 1995; **60**: 16 962–8.
26. National Technical Information Service (NTIS), 5285 Port Royal Road, Springfield, VA 22161, USA. Tel. (703) 487–4650.
27. Higher mortality risk with 100-mg dose of Manoplex. *FDA Med Bull* 1993; **23**: 3.
28. DuMouchel W. Bayesian data-mining in large frequency tables, with an application to the FDA Spontaneous Reporting System. *Am Statistician* 1999; **53**: 177–90.
29. Fletcher AP. Spontaneous adverse drug reaction reporting vs event monitoring: a comparison. *J R Soc Med* 1991; **84**: 341–4.
30. Finney DJ. The detection of adverse reactions to therapeutic drugs. *Stat Med* 1982; **1**: 153–61.
31. Strom BL, Tugwell P. Pharmacoepidemiology current status, prospects, and problems. *Ann Intern Med* 1990; **113**: 179–81.
32. Finney DJ. Statistical aspects of monitoring for dangers in drug therapy. *Methods Information Med* 1971; **10**: 1–8.
33. Standardization of definitions and criteria of causality assessment of adverse drug reactions: drug-induced liver disorders: report of an international consensus meeting. *Int J Clin Pharmacol Ther Toxicol* 1990; **28**: 317–22.
34. Auriche M, Loupi E. Does proof of causality ever exist in pharmacovigilance? *Drug Safety* 1993; **9**: 230–5.
35. Begaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? *Br J Clin Pharmacol* 1994; **38**: 401–4.
36. Temple RJ, Jones JK, Crout JR. Adverse effects of newly marketed drugs. *New Engl J Med* 1979; **300**: 1046–7.
37. Rawlins MD. Pharmacovigilance: paradise lost, regained or postponed? The William Withering Lecture 1994. *J R College Physicians Lond* 1995; **29**: 41–9.
38. Koch-Weser J, Sellers EM, Zacest R. The ambiguity of adverse drug reactions. *Eur J Clin Pharmacol* 1997; **11**: 75–8.
39. Green DM. Pre-existing conditions, placebo reactions, and “side effects”. *Ann Intern Med* 1964; **60**: 255–65.
40. Reidenberg MM, Lowenthal DT. Adverse nondrug reactions. *New Engl J Med* 1968; **279**: 678–9.
41. Karch FE, Smith CL, Kerzner B, Mazzullo JM, Weintraub M, Lasagna L. Adverse drug reactions—a matter of opinion. *Clin Pharmacol Ther* 1976; **19**: 489–92.
42. Scott HD, Rosenbaum SE, Waters WJ, Colt AM, Andrews LG, Juergens JP, *et al*. Rhode Island physicians’ recognition and reporting of adverse drug reactions. *R I Med J* 1987; **70**: 311–6.
43. Sachs RM, Bortnichak EA. An evaluation of spontaneous adverse drug reaction monitoring systems. *Am J Med* 1986; **81**: 49–55.
44. Weber JCP. *Epidemiology of adverse reactions to nonsteroidal antiinflammatory drugs*. In: Rainsford KD, Velo GP, eds., *Advances in Inflammation Research*, Vol. 6, New York: Raven Press. 1984; 1–7.
45. Serradell J, Bjornson DC, Harzema AG. Drug utilization study methodologies; national and inter-

- national perspectives. *Drug Intell Clin Pharm* 1987; **21**: 994–1001.
46. Center for Drug Evaluation and Research, Food and Drug Administration. *Annual Adverse Drug Experience Report 1996*. Rockville, MD: CDER, 1996.
47. Rossi AC, Knapp DE. Discovery of new adverse drug reactions: a review of the Food and Drug Administration's Spontaneous Reporting System. *J Am Med Assoc* 1984; **252**: 1030–3.
48. Milstien JB, Faich Ga, Hsu JP, Knapp DE, Baum C, Dreis MW. Factors affecting physician reporting of adverse drug reactions. *Drug Info J* 1986; **20**: 157–64.
49. Sekul EA, Cupler EJ, Dalakas MC. Aseptic meningitis associated with high-dose intravenous immunoglobulin therapy frequency and risk factors. *Ann Intern Med* 1994; **121**: 259–62.
50. Scribner CL, Kapit RM, Phillips ET, Rickles NM. Aseptic meningitis and intravenous immunoglobulin therapy. *Ann Intern Med* 1994; **121**: 305–6.
51. Blum MD, Graham DJ, McCloskey CA. Temafloxacin syndrome review of 95 cases. *Clin Infect Dis* 1994; **18**: 946–50.
52. Hertzman, PA. L-tryptophan related eosinophilia–myalgia syndrome. In: *Drug and Device Induced Disease: Developing a Blueprint for the Future. Proceedings of a MedWatch Conference 1994, Washington, DC*. Rockville, MD: FDA, 1994; 20–2.
53. Love LA, Rader JI, Crofford LJ, Raybourne RB, Principato MA, Page SW, *et al*. Pathological and immunological effects of ingesting L-tryptophan and 1,1'-ethyldenebis (L-tryptophan) in Lewis rats. *J Clin Invest* 1993; **91**: 804–11.
54. Farinelli S, Mariani A, Grimaldi A, Mariani M, Iannesi A, De Rosa F. Eosinophilia–myalgia syndrome associated with 5-OH-tryptophan. Description of a case. *Recenti Prog Med* 1991; **82**: 381–4.
55. Lampert A, Joly P, Thomine E, Ortoli JC, Lauret P. Scleroderma-like syndrome with bullous morphea during treatment with 5-hydroxytryptophan, carbidopa and flunitrazepam. *Ann Dermatol Venereol* 1992; **119**: 209–11.
56. Flockhart DA, Clauw DJ, Sale EB, Hewett J, Woosley RL. Pharmacogenetic characteristics of eosinophilia–myalgia syndrome. *Clin Pharmacol Ther* 1994; **56**: 398–405.
57. Okada S, Sullivan EA, Kamb ML, Philen RM, Miller FW, Love LA. Human leukocyte antigen (HLA) DRB1 alleles define risk factors for the development of eosinophilia myalgia syndrome. *Arthritis Rheum* 1994; **37**: S274.

11

Spontaneous Reporting Systems Outside the US

BENGT-ERIK WIHOLM

Department of Clinical Pharmacology, Karolinska Institutet, Stockholm, Sweden

STEN OLSSON

The WHO Collaborating Center for International Drug Monitoring, Uppsala, Sweden

NICHOLAS MOORE

Department of Pharmacology, Université Victor Segalen, 33076 Bordeaux, France

PATRICK WALLER

Medicines Control Agency, London, UK.

INTRODUCTION

The broad awareness that modern drugs could carry unexpected hazards was triggered by a letter to the editor of *The Lancet* published on 16 December 1961.¹ In this historical document of 15 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, the governments in many countries arranged procedures for a systematic collection of information about suspected adverse drug reactions (ADRs). These systems were based on the spontaneous reporting of suspected ADRs by physicians and were first organized in Australia, Canada,

Czechoslovakia, Ireland, the Netherlands, New Zealand, Sweden, the UK, the US, and West Germany. Each of these was initiated between 1961 and 1965. Similar systems now operate in more than 60 countries.

In 1968, ten countries from Australasia, Europe, and North America agreed to pool all reports which had been provided 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 WHO scheme was set up at WHO headquarters in Geneva in 1970. The economic and operative responsibilities of the WHO Center were transferred to Sweden in 1978. The formal responsibility for and the coordination of the program, however, still rest with WHO headquarters in Geneva. Today 56

countries participate in the program as full members and a further seven countries^{2,3} as associate members (Figure 11.1).

Spontaneous reporting systems are primarily designed to detect new ADRs: they generate signals about possible ADRs, creating hypotheses for subsequent 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; and to identify special risk groups. Under certain selected circumstances, spontaneous reporting can be used to provide valuable information for these latter tasks, as well.

In this chapter we intend to describe the organization, operation, and results from spontaneous reporting schemes outside of the US.

DESCRIPTION

ADR reporting schemes differ in a number of dimensions. First, there are in essence two parallel, more or less global systems.

1. *The medical literature*: many journals publish case reports of patients who experienced 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 system rather than the first. Literature case reports are, however, included in some of the national systems. In the second category there are now two

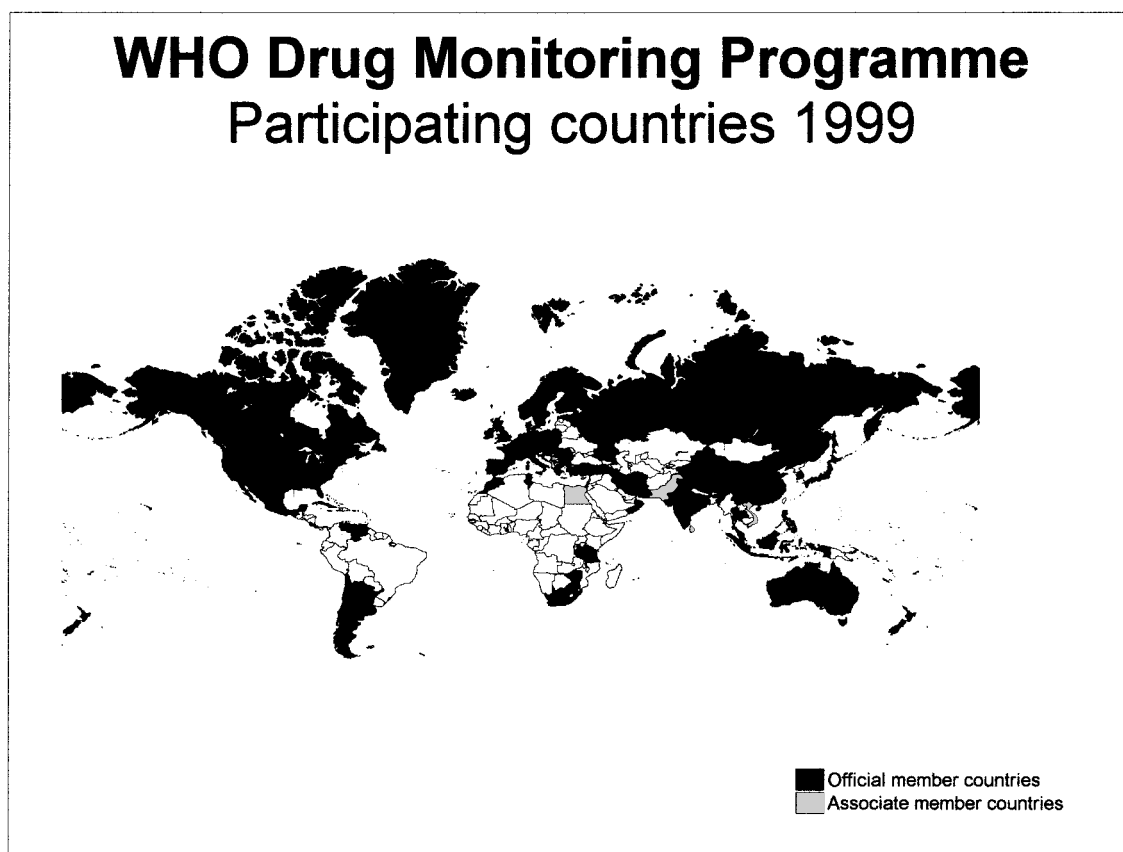


Figure 11.1. A map showing countries participating in the Drug Monitoring Program.

international systems, one under the auspices of WHO in which data on all suspected ADRs are pooled and coordinated by the Uppsala Monitoring Centre in Sweden, and the European Union (EU) pharmacovigilance system. In the latter, all member states and the European Medicines Evaluation Agency (EMA) are connected via secure intranet (Eudranet) for the exchange of pharmacovigilance information. A database, Eudrawatch, for the collation and analysis of reports of serious ADRs associated with products authorized through the EU centralized procedure, is also under development.

National systems themselves are organized in many different ways.³ 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 the pharmaceutical manufacturers, including the largest national system, that of the US (see Chapter 10). Most centers review each report on an individual basis using a clinical pharmacology approach, often making judgements about each case as to how likely it is that the drug caused the adverse event (see also Chapter 32). However, others use mainly an aggregate or epidemiological approach to the analysis of the reports (see Chapter 10). Finally, the national centers differ dramatically in how they interact with reporters. 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 dialogue between the reporters and the center.

Each of these differences will be described in more detail.

ORGANIZATION, AFFILIATION, AND TASKS OF NATIONAL MONITORING CENTERS

In most countries, the monitoring center is part of the drug regulatory authority. In some countries, e.g., The Philippines and New Zealand, the functions are carried out jointly by the drug

control authority and a university institution. 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.

Also, in France a new drug control agency has been formed. The French Medicines Agency, formed in 1994, has taken up duties formerly part of 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, reported to and agreed 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, result in delays in the flow of information.

In Morocco, New Zealand, and Tanzania, the national centers also function as Poison Information Centers. These may serve as useful models for other countries, since intoxication and adverse reactions are often related. The regional centers in France and in Sweden also have responsibility for a general drug information activity, providing

pharmacological advice for specific patients. 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 they receive clinically relevant information from the center.

Some regional centers, e.g., in Barcelona in Spain, Bordeaux in France, and those in Sweden, are also engaged in formal pharmacoepidemiology studies to follow up potential signals created by the spontaneous reports.

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 and 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. In Denmark and Sweden, a list of drugs of special interest is published in the national medical journal. In New Zealand and Ireland, some selected new drugs are put in an Intensive Reporting program. Most countries have, however, issued rather general recommendations as to what type of reactions should be reported to the National Center.

- In at least ten countries, including France, Norway, and Sweden, it is mandatory by law for physicians and dentists to report cases of suspected serious adverse reactions to the regulatory authority. France has also written "Good Pharmacovigilance Practices" which set rules for the reporters (including access to raw clinical data), for the regional centers, for the pharmaceutical companies, and for the Agency in the proper management of individual cases, and case series.
- In some 25 countries, including the EU, Japan, and the US, it is obligatory for pharmaceutical

companies (MAHs in the EU) to submit to the regulatory authority cases of suspected adverse experiences that have become known to them.

SOURCES OF REPORTS

The regulatory status and the 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; and
- 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 it is normally read by a physician or a pharmacist, who makes a judgment about whether the information provided is sufficient as a basis for an opinion on the correctness of the diagnosis and causality, or whether more data should be requested. In some countries, pharmacists who have access to medical consultants staff the national centers. 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 the national center is aided in making the final assessment of causality and the evaluation of the clinical

importance of the aggregate reports by an advisory committee of experienced clinical experts.

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. This effort was started in France, and then international consensus was reached in meetings under the auspices of CIOMS (Council of International Organizations of Medical Sciences, which is linked to the WHO). For example, internationally agreed upon criteria and definitions have been published for hematological and hepatic reactions^{4,5} and some other types of reactions.⁶⁻⁹

No common standard for the detailed operational assessment of causality has been agreed upon internationally. Most experts are able to 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 an ongoing scientific debate (see Chapter 32). During the last 15 years, a number of more or less complicated and comprehensive algorithms for the assessment of causality have been constructed.^{10,11} When tested by their inventors, these algorithms have, in general, been found to decrease inter-rater variability.¹²⁻¹⁴ This has not, however, always been the case when independent groups^{15,16} have

tested the algorithms. Moreover, it has not been possible to test whether the assessments reached by the use of the 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 that today is 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.¹⁷ The use of this algorithm is also mandatory for the pharmaceutical industry. Beyond causality, the major interests of having a common assessment method are the harmonization of data content of case reports, and the didactic value of the exercise. In contrast, 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. 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 causality in a semi-quantitative way (Table 11.1). Methods for

Table 11.1. 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 (dechallenge) should be clinically plausible. The event must be definitive pharmacologically or phenomenologically, using a satisfactory rechallenge 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 (dechallenge). Rechallenge 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.

Unassessible/unclassifiable. A report suggesting an adverse reaction that cannot be judged because information is insufficient or contradictory, and which cannot be supplemented or verified.

assessing causality in case reports are discussed in more detail in Chapter 32. 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.

FEEDBACK TO REPORTERS

Some form of feedback from the national center has to be arranged for clinicians to be involved in two-way communication. In many countries, each reporter receives a personal acknowledgement, often including a preliminary evaluation of the case. In France, feedback is left to the initiative of the individual regional centers. It is usually done on a case-by-case basis, sometimes with regularly produced information bulletins.

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.

GENERATION AND EVALUATION OF SIGNALS

Spontaneous adverse drug reaction reporting is principally a method of identifying previously unrecognized hazards with marketed medicines. The process of generating “signals” of such hazards from the data has been likened to looking for a needle in a haystack.¹⁸

In the early days, when reports were relatively few, signals were looked for manually or through checking, for example, quarterly lists. 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.^{19–21} Later, differences in such proportions were tested by statistical significance tests. A signal published based on such a test was the higher proportion of serum-sickness-like reactions to cefaclor, than to other cephalosporins and ampicillin.²²

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.²³ Using this technique, a strong signal was found for an association between ACE inhibitors and hypoglycemia. This signal disappeared, however, when the use of antidiabetics was controlled for in the analysis. These analyses were done manually, often on a case-by-case basis.

The WHO Collaborating Center, on the basis of reports received from member countries, increasingly produces signals of potential but previously unknown adverse reactions. Each quarter more than 2000 new potential drug–reaction combinations are identified. The WHO Center has created a system whereby independent experts in an international review panel screen these new signals. The results of these screenings are then circulated among the countries. Several such signals have been published, based on single or a few cases that appear in two or more countries and where the material was pooled to generate a signal of better quality (see Table 11.2). Compilations of all ADRs reported to a drug or a drug class are also published in medical journals or books.^{19–22} Very recently, computer-aided systems for the identification of new potential signals have been created in the UK and by the Uppsala Monitoring Centre.

Proportional Reporting Ratios

The UK system is a statistical aid to signal generation that is based on a proportionate and comparative approach. The proportion of all reactions to a drug which are a particular medical condition of interest is compared to the same proportion for all drugs in the database. The resulting statistic is called a proportional reporting ratio (PRR). Judgements about signals may then be made using the PRR, along with the associated value of χ^2 and the absolute number of reports.

This approach uses the total number of reports for the drug as a denominator to calculate the proportion of all reactions which are the type of interest (e.g., hepatitis). This proportion may be compared with the value for other drugs. It is also

Table 11.2. Examples of publications on new adverse drug reactions during the 1980s and 1990s that have been discovered through spontaneous reports or where spontaneous reports have reinforced previous suspicions

Reaction	Drug	Year of publication	Country	Ref
Agranulocytosis	Mebhydrolin	1982	Australia	53
Diplopia	β -blockers	1982	UK	54
Testicular pain	Mazindol	1983	Australia	55
			Netherlands	
Drug interaction	Oral Contraceptive/Griseofulvin	1984	Netherlands	56
Fibrosis	Ergotamine	1984	Belgium	57
Photosensitivity	Azapropazone	1985	6 countries	58
Acute hypersensitivity	Paracetamol	1985	10 countries	59
Agranulocytosis	Indalpine	1985	France	60
Extrapyramidal disorders	Flunarazine/Cinnarazine	1986	Spain	61
Skin reactions	Terfenadine	1986	5 countries	62
Esophageal obstruction	Glucomannan	1986	Australia	63
Pulmonary infiltrates	Tolfenamic acid	1987	Finland	64
Angioedema	Enalapril	1987	UK	65
Urticaria	Captopril	1988	Netherlands	66

possible to compare complete profiles of ADR reporting for drugs of different types of reaction, where differences in the profile may represent signals. The result of such a calculation is called a proportional reporting ratio (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 expected or null value for a PRR is 1.0 and the numbers generated are measures of association that behave in fashion similar to relative risks. The higher the PRR, the greater the strength of the signal. Measures of statistical association for each value are calculated using standard methods such as 95% confidence intervals and Fisher's exact test or the χ^2 test with one degree of freedom (with Yates's correction). Judgment about whether or not there is a signal, and its strength, may then be made on the basis of

three key pieces of information, i.e., the PRR, the value of χ^2 , and the absolute number of reports. A minimum signal would be a PRR of at least 2, χ^2 of at least 4, and three or more cases.

PRRs have some advantages over calculation of reporting rates. First, no external data are needed and the limitations of such data (including delay in receipt) do not apply. Second, 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 that are of a specified type.

It is 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 χ^2 values are measures of association and not causality.

There are a number of possible extensions to the method that are being evaluated further. For example, PRR calculations could be restricted to particular groups of drugs, to serious or fatal reports, or to particular age groups. Examination of changes in PRRs over time may help to demonstrate how signals of adverse drug reactions can be identified as early as possible.

A Bayesian Approach Using Neural Networks

RATE OF REPORTING

The system developed by the Uppsala Monitoring Centre is based on Bayesian statistics and the use of a neural network computer program.²⁴ Every quarter the reports that have come in are compared with those already in the database for proportions of specific drug–reaction combinations. If a specific drug–reaction combination appears significantly more often in the new reports than before, a signal is automatically generated.

Figure 11.2 depicts the number of reports received by the national centers in 1994.² In most countries reporting has gradually increased over time. It usually takes some five to ten years of operation before reporting reaches a stable level. The number of reports has, however, increased substantially in Australia, Spain, and the UK. The number of reports relayed to the WHO Center is often less than that received in the country, for various technical reasons. In France,

CATEGORIES SUBMITTING REPORTS

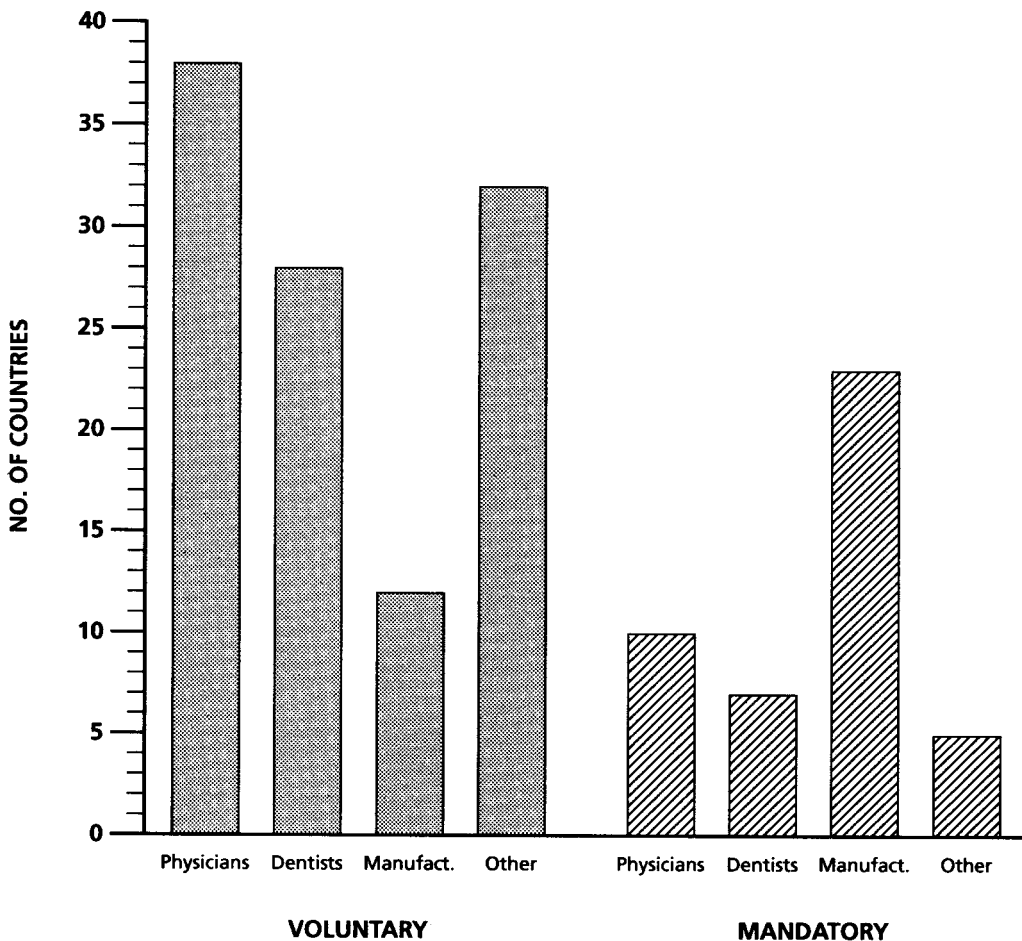


Figure 11.2. Number of case reports submitted to the WHO Collaborating Center each year.

reports received by the national agency from the manufacturers (50%) are not sent to the WHO, 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 10).

INTERNATIONAL REPORTING

In recent years there has been an increasing tendency among regulators to demand that manufacturers report directly to them suspected ADRs that happened in other countries, in spite of the fact that most of these 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 the 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 the Council of International Medical Organizations (CIOMS), which is affiliated to the WHO. This initiative, called “CIOMS I,”²⁵ is now accepted in most countries and by most manufacturers, and has been accepted by the International Conference on Harmonization (ICH) as a guideline. This system has definitely decreased the diversity of the rules regarding international reporting. 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 this CIOMS I initiative produced. They preferred periodic safety updates, including the evaluation of the safety situation at large by the company. In some countries laws and regulations mandated such safety updates, but again there were differences in the rules, formats, and time schedules for such safety updates. Therefore, a second CIOMS initiative—“CIOMS II”—was undertaken to harmonize the contents, format, and time schedule for periodic safety updates.²⁶ 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 postmarketing studies, and the principle that the manufacturer should write an overall safety evaluation of the product.

The CIOMS II safety updates are now unofficially accepted by the ICH 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 ICH format.

The European Parliament has adopted a regulation creating the European Agency for the Evaluation of Medicinal Products (EMA), which is located in London, UK. According to these regulations, all serious suspected adverse reactions that are reported to a Marketing Authorization Holder (MAH) by a health professional shall be reported to the health authority of the country in which it occurred within 15 days. Health authorities shall also report to the EMA and to the MAH within 15 days. MAHs are also mandated to report ADRs occurring outside the EU as well as from the world literature that are both serious and unexpected (unlabeled).

The problem of managing and distributing large numbers of ADR reports within 15 days has led to a commitment from many to communicate data electronically. However, standardization of data sets and terminology is required for this to be implemented effectively. The Medical Dictionary of Regulatory Activities (MedDRA)²⁷ and electronic standards for the transfer of regulatory information were introduced to the ICH (M1 and M2 topics) in 1994. Both are applicable to pre- and postmarketing phases of the regulatory cycle. MedDRA was based on the terminology developed by the Medicines Control Agency in the UK. It provides greater specificity of data entry (more terms) 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 small.

STRENGTHS

A spontaneous reporting system can be relatively inexpensive to operate. One physician, one pharmacist, and a secretary can usually manage between one and two thousand reports a year, depending on the amount of scrutiny, followup, feedback, and other activities that are part of the program. The basic technical equipment needed is also a relatively minor investment. Together with other properties (Table 11.3), some of which are unique, this makes a spontaneous reporting system one of the 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 is not restricted to either hospitalized patients or those treated as outpatients, and does not exclude patients treated for other concomitant diseases etc. 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, such as the elderly, or in combination with other drugs.

In an early analysis of how important ADRs were first suspected and then verified, Venning²⁸ found that 13 out of 18 reactions were first signaled

Table 11.3. Strengths and weaknesses 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 followup studies of patients with severe ADRs, to study mechanisms

Weaknesses

The amount of clinical information available is often too limited to permit a thorough case evaluation
 Under-reporting decreases sensitivity and makes the systems sensitive to selective reporting
 The reporting rate is seldom stable over time
 No direct information on incidence

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.^{29,30} The opposite situation was found in only a few cases. Some of these signals were published. Table 11.2 presents these and some other examples of situations where a drug problem has been “discovered” or where a previous signal has been reinforced by the use of spontaneous reports. 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.

WEAKNESSES

Spontaneous reporting systems are mainly intended to produce signals about potential new ADRs. To fulfill this function properly, one must recognize that a number of false signals will be produced and, therefore, 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 epidemiological study design and one or more of the data resources described in Chapters 12–25.

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 the signs and/or symptoms of his patient may be caused by a drug. 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, e.g., the first cases of the oculomucocutaneous syndrome, Guillain–Barré syndrome, 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, there are a number of more or less irrational reasons for not reporting this suspicion. Ignorance of the value of ADR reporting, ignorance of the reporting rules, and too little time to report have been given as reasons for not reporting. Increased information and feedback by national agencies, the medical schools, the pharmaceutical manufacturer, and professional journals, in cooperation, could rectify these inadequacies. The health care providers also have a clear role here. It is their responsibility to monitor the quality of health care and to build in ADR reporting practices in their quality assurance systems, as well as in continuing medical education.

Besides delaying the detection of new ADRs, under-reporting 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 a minimum level. More important is that under-reporting 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. For example, in Sweden, as in many other countries, there was a surge of reports of gastrointestinal ulcers and bleeding suspected to have been caused by piroxicam soon after its introduction. If the number of ulcers and bleeding reported for piroxicam is compared with that reported for

indomethacin or aspirin, one may be led to believe that the new drug was much more ulcerogenic than the old ones. However, there are many other possible reasons for the 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. 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 fourfold difference in reports of hemorrhagic diarrheas in relation to sales of the two penicillin-v products Calciopen® and Kåvepenin® that was discovered some years ago in Sweden. An analysis of the situation failed to reveal any differences in the two products. Actually, 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 by private practitioners, groups who traditionally do not report ADRs. This could not, however, totally explain the apparent difference in ADR rate, as the reporters seldom were the prescribers. The most important explanation was probably that the younger product was more commonly recommended and used in “high reporting” counties.

PARTICULAR APPLICATIONS

WITHOUT THE ADDITION OF OTHER DATA

A spontaneous reporting system can, in its basic form, be regarded as an incomplete (and at worst biased) case series, without any information on the size or characteristics of the population exposed to the drug. In this basic form it is rarely possible to use spontaneous reports to establish a causal connection between an adverse event and a drug, unless:

- (a) there is at least one case with a positive rechallenge and some other supportive cases

who do not have known confounding drugs or diseases, or

- (b) there is a cluster of exposed cases reported, the background incidence of the adverse event is close to zero, and there is no confounding.

The reappearance of an adverse event when a drug is given again is certainly no proof of causality.³¹ In practice, however, one is rather reassured that there is strong evidence for a causal connection 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. Of course, this is only possible if the medical event in question were 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 infections³² will convince most clinicians that this drug did and can cause such a reaction. For typical “hit and run” effects like thromboembolic diseases and for diseases that can be cyclic, information on rechallenge, however, can be misleading. For example, a young boy in Sweden 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.

Information on rechallenge is relatively uncommon in most spontaneous reporting systems. Planned rechallenge 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.³³ In the French pharmacovigilance database, re-exposure is reported in 8.5% of reports, positive in 6%.³⁴

An example meeting the second criterion occurred with the cardiovascular drug aprindine. Four to five cases of agranulocytosis were reported in the Netherlands during the first two years the drug was marketed. As the background incidence of agranulocytosis is only five to eight per million

inhabitants per year,³⁵ this made a strong case for a causal relationship.³⁶

WITH THE ADDITION OF DENOMINATOR DATA

Today, most drug regulatory authorities in industrialized countries and pharmaceutical manufacturers have access to 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 29). In many countries information on drug sales and prescribing is confidential, but in the Nordic countries this information is published periodically.³⁷ Another source for information on pharmaceutical sales and prescribing habits in a large number of countries is Intercontinental Medical Statistics International (IMS). Data in a number of countries from IMS have been combined with ADR information from the WHO database to provide rough incidence estimates of certain drug-induced problems.^{38,39}

WITH THE ADDITION OF NUMERATOR DATA

If the rate of reporting is known, the estimate of the numerator can become more accurate. From studies using registers of hospital discharge diagnoses, it has been possible to calculate reporting rates for some areas, ADRs, and periods of time. Considering serious reactions such as blood dyscrasias, thromboembolic disease, and Stevens–Johnson syndrome, in general between 20% and 40% of the patients discharged with these diagnoses have been found to be reported.⁴⁰ 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.⁴¹ However, reporting rates probably cannot be generalized. The magnitude of under-reporting is important to know when evaluating the data, but should not be used to correct for under-reporting in the calcula-

tions since the reporting rate is time-, problem-, drug-, and country-specific.

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 often possible to derive a rough estimate of the frequency or incidence rate of an ADR. Such estimates can, of course, 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. For very rare reactions, they may actually be the only conceivable measure.

Specifically, with knowledge of the number of defined daily doses (DDDs) sold and the average prescribed daily dose (PDD), it is possible to obtain 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 rough estimate of the incidence.⁴² If prescription statistics are available, the number of prescriptions may actually be a better estimate of drug use among outpatients than the number of treatment weeks calculated from sales data, especially for antibiotics, where drug use is mostly short term and doses and treatment times may vary with patient age and indication. As an example, the frequency of reports of serum-sickness-like reactions and erythema multiforme was 17 and four per 10 000 prescriptions, respectively, among children aged 0–9 years prescribed a solution of cephaclo (a cephalosporin antibiotic). No such reactions were reported among adults using tablets. These results could imply that there was something in the solution apart from the active substance that caused the reactions. However, other possible explanations are age-dependent differences in reporting or in actual immunological reactivity, and one must always be extremely careful in the interpretation of such data.

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 that is used mainly for the treatment of glaucoma) have been reported since the drug was introduced in the mid-1950s.⁴³ There are no estimates of the incidence of this reaction, but it was certainly thought to be very rare, probably more rare than aplastic anemia occurring after the use of chloramphenicol. Between 1972 and 1988, 11 cases were reported to have occurred in Sweden.⁴⁴ 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, giving a reported incidence of about one in 20 000, or 50 per million patient years. From a population based case-control study of aplastic anemia in which Sweden participated,⁴⁵ it could be estimated that the total yearly incidence of aplastic anemia in the relevant age groups was about six per million. 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 people 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 under-reporting in this example is unknown. However, in one study the reporting rate for aplastic anemia was found to be 30%,⁴⁶ 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 which, on clinical pharmacological grounds, seemed to be reasonable alternatives. It is a clear limitation that multiple drug exposures cannot be corrected for in a rough analysis such as this.

In March 1982 a new antidepressant drug, zimeldine, was introduced in Sweden. Zimeldine

was a new selective serotonin uptake blocker. A hypersensitivity reaction that mimicked influenza—with fever, myalgia, arthralgia, and sometimes a slight rash or elevation of the liver transaminases—seemed to be the only characteristic new ADRs associated with the use of this new compound.

However, during the first 16 months of the marketing of zimeldine, 20 patients were reported to have developed various neurological complications during treatment with the drug. Eight of these patients fulfilled the criteria for suffering from Guillain-Barré syndrome.⁴⁷ All of these patients developed a “flu-like” hypersensitivity reaction within the first month of treatment. This neurological complication developed in close connection to the hypersensitivity reaction. By using sales and prescription statistics a total exposure time of about 14 000 patient years was estimated. The average treatment time was 60–100 days and a maximum of 60 000 patients were estimated to have been treated with zimeldine. From Sweden’s hospital statistics, the yearly incidence of Guillain-Barré syndrome was estimated to be 2.5/100 000 inhabitants. The reported frequency of the Guillain-Barré syndrome among patients exposed to zimeldine could then be estimated to be at least about 1/6000, the incidence rate 0.6/1000 treatment years, the relative risk 24, and the excess risk 0.58/1000 treatment years.

In this situation no relevant confounding factors could be identified. All reported cases fulfilled strict criteria for the disease, so misclassification of the cases could be ruled out. Complete ascertainment of all cases (complete reporting) could not be guaranteed. However, this would lead to an underestimate of the risk estimates. Moreover, the population at risk (the denominator) might have been grossly overestimated. All patients who developed the Guillain-Barré syndrome did so between day 11 and day 30 of treatment. Thus, the risk did not seem to be independent of time. However, the denominator was calculated from the estimate of the total patient exposure time, which included patients who stopped treatment within the first week as well as those treated for several months.

In some instances it has been possible to compare risk estimates from a formal epidemiolo-

gic case-control study with those derived from the Swedish drug monitoring system. The relative risks for agranulocytosis induced by co-trimoxazole and sulfasalazine were astonishingly alike.⁴⁸

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, whether any group of patients is at a particularly increased risk, and whether 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 hypoglycemic episodes 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.⁴⁹

In one of the first followup 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 of the blood group O more often than would have been expected from the distribution of blood groups in the population.⁵⁰

A similar study⁵¹ investigated patients reported to have developed lupoid reactions while taking hydralazine for hypertension. A much higher percentage was slow acetylators than the 40% expected from the distribution of this phenotype in the population at large.

Finally, in a more sophisticated study, Strom 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.⁵² 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.

THE FUTURE

At least in the 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 as to type 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 programs. Moreover, it would be naive to assume that we will not be confronted with totally new and unexpected types of ADRs in the years to come. Thus, the importance of postmarketing surveillance will not diminish; we must continuously develop our total armamentarium of methods for this task. Spontaneous reporting of new and unexpected reactions is likely to remain one of the basic methods for pharmacoepidemiology for many years.

Pharmacovigilance programmes are now being established also in developing countries from which we have not had much information in the past. It is likely that monitoring of populations with another pattern of morbidity and a different nutritional status will reveal different types of adverse reactions from established medicines than what we have learned to expect from populations in the industrialized world. Influence of co-

medication with traditional medicines and unexpected failure of efficacy because of substandard or counterfeit medicines will have to be covered by the pharmacovigilance systems.

The role of spontaneous reporting in the future will be even more central if it can be developed further. The basic requisite for its enhanced effectiveness is an increased flow of information, in both 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 automated transfer of computerized hospital discharge diagnoses, could be used to study case series.

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 mandatory to enhance the practicing clinician's awareness of and cooperation with ADR reporting. Here a regional system with mutual benefits, such as the French system, seems promising.

Periodically it may be beneficial to focus on certain new drug classes to clarify their ADR spectrum as soon as possible, e.g., the HIV ADR reporting scheme in the UK.

Finally, it is not enough to increase the reporting signaling function only quantitatively. To gain new knowledge and to guarantee drug safety, it is necessary to increase the qualitative aspects of the information. The number of signals of possible new adverse events that are reported (2000/quarter to WHO) calls for a more efficient procedure for scrutiny and evaluation than can be accomplished with the present resources in the WHO program. We must therefore set up a system where these signals can be analyzed and evaluated more effectively, such as the proportional reporting ratios in the UK and the neural network system developed by the Uppsala Monitoring Centre. This can only be accomplished through

increased resources for and a more effective collaboration among national monitoring centers, university institutions, and particularly pharmaceutical manufacturers.

DISCLAIMER

These are our views and not necessarily those of our respective employers.

REFERENCES

1. McBride WG. Thalidomide and congenital abnormalities. *Lancet* 1961; **2**: 1358.
2. Olsson S. The Role of the WHO Programme on international drug monitoring in coordinating worldwide drug safety efforts. *Drug Safety* 1998; **19**: 1–10.
3. Olsson S, ed. *National Pharmacovigilance Systems*, 2nd edn. Uppsala: Uppsala Monitoring Centre, 1998.
4. Benichou C. Criteria of drug-induced liver disorders. Report of an International consensus meeting. *J Hepatol* 1990; **11**: 272–6.
5. Benichou C, Solal-Celigny P. Standardization of definitions and criteria for causality assessment of adverse drug reactions. Drug-induced blood cytopenias: report of an international consensus meeting. *Nouv Rev Fr Hematol* 1991; **33**: 257–62.
6. CIOMS. Basic requirements for the use of terms for reporting of drug reactions. *Pharmacol Drug Safety* 1992; **1**: 39–45.
7. CIOMS. Basic requirements for the use of terms for reporting of drug reactions (II). *Pharmacol Drug Safety* 1992; **1**: 133–7.
8. CIOMS. Harmonizing the use of adverse drug reactions and minimum requirements for their use: respiratory disorders and skin disorders. *Pharmacoepidemiol Drug Safety*; 1997; **6**: 115–27.
9. CIOMS. Basic requirements for the use terms for reporting adverse drug reactions (VIII): renal and urinary system disorders use. *Pharmacoepidemiol Drug Safety*. 1997; **6**: 203–11.
10. Venulet J, Berneker G-C, Ciucci AG, eds. *Assessing Causes of Adverse Drug Reactions*. London: Academic Press, 1982.
11. Herman RL. Drug–event association: perspectives, methods, and uses. *Drug Inf J* 1984; **18**: 195–337.
12. Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. *Clin Pharmacol Ther* 1977; **21**: 247–54.
13. Kramer MS, Leventhal JM, Hutchinson TA, *et al*. An algorithm for the operational assessment of adverse drug reactions. *J Am Med Assoc* 1979; **242**: 623–32.
14. Naranjo C, Busto EM, Seller P. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; **30**: 239–45.
15. Louik C, Lacouture PG, Mitchell AA, *et al*. A study of adverse reaction algorithms in a drug surveillance program. *Clin Pharmacol Ther* 1985; **38**: 183–7.
16. Pere JC, Begaud B, Haramburu F, *et al*. Computerized comparison of six adverse drug reaction assessment procedures. *Clin Pharmacol Ther* 1986; **40**: 451–61.
17. Moore N, Biour M, Paux G, *et al*. Adverse drug reaction monitoring: Doing it the French way. *Lancet* 1985; **2**: 1056–8.
18. Edwards IR. Adverse drug reactions: finding the needle in the haystack. *Br Med J* 1997; **315**: 500.
19. Holmberg L, Boman G, Bottiger LE, *et al*. Adverse reactions to nitrofurantoin. Analysis of 921 reports. *Am J Med* 1980; **69**: 733–8.
20. Weber JCP. Epidemiology in the United Kingdom of adverse drug reactions from nonsteroidal anti-inflammatory drugs. In: Rainsford KD and Velo GP, eds, *Side Effects of Anti-Inflammatory Drugs*. Lancaster: MTP Press, 1986, Part I, 27–36.
21. Wiholm B-E, Myrhed M, Ekman E. Trends and patterns in adverse drug reactions to nonsteroidal anti-inflammatory drugs reported in Sweden. In: Rainsford KD and Velo GP, eds, *Side Effects of Anti-Inflammatory Drugs*. Lancaster: MTP Press, 1986, Part I, 55–62.
22. Stricker BHCh, Tjissen JGP. Serum-sickness like reactions to cephachlor. *J Clin Epidemiol* 1992; **45**: 1177–84.
23. Moore N, Kreft-Jais K, Haramburu F, *et al*. Reports of hypoglycaemia associated with the use of ACE inhibitors and other drugs: a case/non-case study in the French Pharmacovigilance System database. *Br J Clin Pharmacol* 1997; **44**: 513–8.
24. Bate A, Lindqvist M, Edwards IR, Olsson S, Orre R, Lansner A, De Freitas RM. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol* 1998; **54**: 315–21.
25. CIOMS Working Group. *International Reporting of Adverse Drug Reactions*. Geneva: CIOMS, 1990.
26. CIOMS Working Group II. *International Reporting of Periodic Drug-Safety Update Summaries*. Geneva: CIOMS, 1992.
27. Wood KL, Wood SM. The international medical terminology for regulatory activities: a tool to improve the utilisation of regulatory data and to support its communication within and between organisations. In: Mitchard M. ed., *Electronic Communication Technologies: a Practical Guide for*

- Healthcare Manufacturers*. Buffalo: Interpharm Press Inc., Buffalo Group.
28. Venning GR. Identifications of adverse reactions to new drugs. II—How were 18 important adverse reactions discovered and with what delay? *Br Med J* 1983; **286**: 289–92.
 29. Wiholm B-E, Lindquist M. The detection and evaluation of drug induced agranulocytosis by spontaneous reports. Abstract, *III World Conference on Clinical Pharmacology and Therapeutics, Stockholm, 1986*.
 30. Li D, Lindquist M, Edwards IR. Evaluation of early signals of drug-induced Stevens–Johnson's Syndrome in the WHO ADR Database. *Pharmacol Drug Safety* 1992; **1**: 11–9.
 31. Rothman KJ. Causal inference in epidemiology. In: *Modern Epidemiology*. Boston: Little, Brown and Company, 1986, 7–21.
 32. Carlson J, Wiholm B-E. Trimethoprim associated aseptic meningitis. *Scand J Infect Dis* 1987; **19**: 687.
 33. Nordic Council on Medicines. *Drug Monitoring in the Nordic Countries. An Evaluation of Similarities and Differences*, NLN Publication No. 25. Uppsala: Nordic Council on Medicines, 1989.
 34. Moore N, Noblet C, Kreft-Jais C, Lagier G, Ollagnier M, Imbs JL. French pharmacovigilance database system: examples of utilization. *Therapie* 1995; **50**: 557–62.
 35. The International Agranulocytosis and Aplastic Anemia Study. Risk of agranulocytosis and aplastic anemia. A first report of their relation to drugs with special reference to analgesics. *J Am Med Assoc* 1986; **256**: 1749–57.
 36. van Leeuwen R, Meyboom RHB. Agranulocytosis and aprinidine. *Lancet* 1976; **2**: 1137.
 37. Nordic Council on Medicines. *Nordic Statistics on Medicines 1975–1977*, Part II. Uppsala: Nordic Council on Medicines, 1978.
 38. Lindquist M, Pettersson M, Edwards IR, *et al*. How does cystitis affect a comparative risk profile of tiaprofenic acid with other non-steroidal antiinflammatory drugs? An international study based on spontaneous reports and drug usage data. *Pharmacol Toxicol* 1997; **80**, 211–7.
 39. Stahl MMS, Lindquist M, Pettersson M, *et al*. Withdrawal reactions with selective serotonin reuptake inhibitors as reported to the WHO-system. *Eur J Clin Pharmacol* 1997; **53**, 163–9.
 40. Wiholm B-E. Spontaneous reporting of ADR. In: *Detection and Prevention of Adverse Drug Reactions, Skandia International Symposia*. Stockholm: Almqvist and Wiksell, 1983; 152–67.
 41. Böttiger M, Romanus V, deVerdier C, *et al*. Osteitis and other complications caused by generalized BCG-itis. Experiences in Sweden. *Acta Paediat Scand* 1982; **71**: 471–8.
 42. Böttiger LE, Boman G, Eklund G, *et al*. Oral contraceptives and thromboembolic disease. Effects of lowering oestrogen content. *Lancet* 1980; **1**: 1097–101.
 43. Fraunfelder FT, Meyer MS, Bagby GC Jr, *et al*. Hematologic reactions to carbonic anhydrase inhibitors. *Am J Ophthalmol* 1985; **100**: 79–80.
 44. Mortimer O, Wiholm B-E. *Acetazolamide—pancytopenia*. Bulletin from the Swedish Adverse Drug Reactions Advisory Committee 46, 1985.
 45. The International Agranulocytosis and Aplastic Anemia Study. Incidence of aplastic anemia. The relevance of diagnostic criteria. *Blood* 1987; **70**: 1716–21.
 46. Böttiger LE, Westerholm B. Drug-induced blood dyscrasias in Sweden. *Br Med J* 1973; **3**: 339–42.
 47. Fagius J, Osterman PO, Siden A, *et al*. Guillain–Barré Syndrome following zimeldine treatment. *J Neurol Neurosurg Psychiatry* 1985; **48**: 65–9.
 48. Keisu M, Ekman E, Wiholm B-E. Comparing risk estimates of sulphonamide-induced agranulocytosis from the Swedish drug monitoring system and a case–control study. *Eur J Clin Pharmacol* 1992; **43**: 211–4.
 49. Asplund K, Wiholm B-E. Glibenclamide-associated hypoglycemia, a report on 57 cases. *Diabetologia* 1983; **24**: 412–7.
 50. Jick H, Slone D, Westerholm B, *et al*. Venous thromboembolic disease and ABO blood type. *Lancet* 1969; **1**: 539–42.
 51. Strandberg I, Boman G, Hassler L, *et al*. Acetylator phenotype in patients with hydralazine-induced lupoid syndrome. *Acta Med Scand* 1976; **200**: 367–71.
 52. 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.
 53. McEwen J, Strickland WJ. Mephydrolin napadisylate. A possible cause of reversible agranulocytosis and neutropenia. *Med J Aust* 1982; **2**: 523–5.
 54. Weber JCP. Beta-adrenoreceptor antagonists and diplopia. *Lancet* 1982; **2**: 826–7.
 55. McEwen J, Meyboom RHB. Testicular pain caused by mazindol. *Br Med J* 1983; **287**: 1763–4.
 56. van Dijke CPH, Weber JCP. Interaction between oral contraceptives and griseofulvin. *Br Med J* 1984; **288**: 1125–6.
 57. Rober H, Derbandrengien JP, Blampain JP, *et al*. Fibrotic process associated with long term ergotamine therapy. *New Engl J Med* 1984; **311**: 601–2.
 58. Olsson S, Biriell C, Boman G. Photosensitivity during treatment with azapropazone. *Br Med J* 1985; **291**: 939.
 59. Sriccker BHC, Meyboom RHB, Lindquist M. Acute hypersensitivity reactions to paracetamol. *Br Med J* 1985; **291**: 938–9.

60. Castot A, Efthymiou M-L. Atteintes hematologiques associées à la prise d'indalpine. *Therapie* 1985; **40**: 337-41.
61. Laporte J-R, Capella D. Useless drugs are not placebos: lessons from flunarizine and cinnarizine. *Lancet* 1986; **2**: 853-4.
62. Stricker BHC, van Dijke CPH, Isaacs AJ, *et al.* Skin reactions to terfenadine. *Br Med J* 1986; **293**: 536.
63. Henry D, Mitchell A, Aylward J, *et al.* Clucoman and risk of oesophageal obstruction. *Br Med J* 1986; **292**: 591-2.
64. Strömberg C, Palva E, Alhava E, Aranko K, Idänpään-Heikkilä J. Pulmonary infiltrations induced by tolfenamic acid. *Lancet* 1987; **2**: 685.
65. Wood S, Mann B, Rawlins M. Angio-oedema and urticaria associated with angiotensin converting enzyme inhibitors. *Br Med J* 1987; **294**: 91-2.
66. Markusse HM, Meyboom RHB. Gynecomastia associated with captopril. *Br Med J* 1988; **296**: 1262-3.

12

Intensive Hospital-based Cohort Studies

KEITH BEARD

Victoria Infirmary, Glasgow, Scotland

DAVID H. LAWSON

Royal Infirmary, Glasgow, Scotland

INTRODUCTION

Until the recognition of chloramphenicol-induced aplastic anemia¹ and thalidomide-associated phocomelia,² adverse drug reaction monitoring was almost entirely anecdotal. However, the thalidomide incident, in particular, led many countries to set up agencies to collect and evaluate information on spontaneously reported adverse drug reactions. These schemes were not without limitations and in 1965 Finney,³ a distinguished Scottish epidemiologist and statistician, advocated a more rigorous approach to searching for adverse drug reactions. He recommended routine recording of all demographic and clinical information on hospitalized patients together with all "events" occurring in these patients. This was to be done regardless of whether any links between drugs and events were made by their physicians. In his opinion, detailed analyses of the resulting information, in effect a series of cohort studies of drug recipients, could lead to the detection of new and previously unsuspected adverse drug effects. At about this time, Leighton Cluff and his colleagues in Baltimore were devel-

oping the first major intensive inpatient drug monitoring program.⁴ Their program concentrated on collecting details of suspected adverse drug reactions, but in some instances involved formal hypothesis testing, performed to establish with more certainty that a suspected reaction was truly an adverse drug effect.⁵

These pioneering proposals and studies stimulated interest in this area and several groups began intensive inpatient monitoring programs to record suspected adverse drug reactions (ADRs), using as data collection monitors either medical staff,⁶ pharmacists,⁷ or nurses.⁸ The last named group has been by far the most successful. It was in 1966 that Hershel Jick and the late Dennis Slone started a pilot scheme to assess the feasibility of continuously monitoring large numbers of medical inpatients. Their goal was to determine the frequency with which these patients experienced acute undesired drug effects during hospitalization. So successful was this scheme that it was gradually expanded to medical wards in six different countries. By 1976, information had been collected from over 50 000 medical inpatients,

allowing much original research on the association between short-term drug exposures and acute ADRs to be carried out. After 1977, the focus of the program was changed to surgical inpatients. This new data resource, although smaller, provided additional information about drug exposure and events occurring in operating rooms.⁹ Small numbers of pediatric¹⁰ and psychiatric patients¹¹ were monitored in a similar manner. All these methods and databases together came to be known as the Boston Collaborative Drug Surveillance Program (BCDSP). A review of its published data gives an impressive idea of the scope of this approach, which has subsequently been adopted and applied by other groups.

DESCRIPTION

The aims of intensive hospital-based cohort studies in pharmacoepidemiology are fourfold: (i) to provide information on overall patterns of drug use in hospitals; (ii) to obtain details of acute adverse events attributable to drugs used in hospitals, and to determine whether particular subgroups of hospitalized patients are at greater risk of experiencing ADRs than others; (iii) to obtain information on the frequency of certain major life threatening events occurring during hospitalization, be they drug related or not; and (iv) to identify associations between pre-hospital drug use and diseases or adverse events causing hospital admissions.

The methods used in this type of study have been described in detail by Jick and his colleagues.¹² Briefly, monitors, usually nurses, are employed to collect routine demographic, social, and medical information from consecutive admissions to the hospital. Details of drug exposures before the hospitalization are obtained from patients by the monitor using a standardized interview conducted shortly after admission. Details of all drug exposures during the hospitalization are also recorded, using standardized self-coding data collection forms. The monitors attend ward rounds, where they gather information concerning undesired or unintended events thought by the attending physician to be causally

related to drug therapy. The degree of certainty that any event is indeed drug related is assessed both by the attending physician and, at a later date, by an independent clinical pharmacologist. In addition to ADR data, information on certain specific events occurring during the hospitalization is recorded routinely, whether or not these events are thought to be due to drug therapy. All such events are major illnesses, and can include sudden death, jaundice, pulmonary embolism, renal failure, gastrointestinal bleeding, convulsions, deafness, and psychosis.

The resulting information is then evaluated for accuracy and completeness before being entered into computer files for detailed analysis. During data entry, several standard tests are applied to the information to ensure validity and internal consistency. Thereafter the information is available for detailed analyses. Routine analyses are then undertaken at regular intervals to search for associations in the data.

STRENGTHS

The main advantage of the intensive hospital-based approach to pharmacoepidemiology studies is the very broad and comprehensive nature of its primary information sources. This approach has made many important contributions to our knowledge about the effects of drugs used in hospitals. Included are studies of drug utilization, descriptive studies of ADRs and their predictors, studies testing hypotheses about ADRs, and studies generating hypotheses about ADRs. Additional information is available about drugs used shortly before the hospitalization. Examples of each of these are given in the section on Particular Applications, below.

Armed with these data, the BCDSP occupied a unique position in the early 1970s. Validated data of the highest quality were scarce at this time, as pharmacoepidemiology was just emerging as a unique scientific discipline. Various aspects of the data collection process also broadened its appeal. The "collaborative" nature of the work allowed international comparisons to be made. The demographic data available allowed the role of variables

such as age and sex to be studied in more detail. Finally, the inclusion of social habits such as alcohol and tobacco consumption allowed study of these potential modifying factors on the incidence of ADRs. Thus, by the mid-1970s an impressive and unique body of information had been accumulated using this approach, analyzed in detail and published in highly reputable, peer reviewed journals.

WEAKNESSES

Intensive hospital based cohort studies have provided clinicians and researchers with accurate reproducible details of acute adverse reactions to virtually all commonly used drugs available in the USA and in Europe in the 1960s and 1970s. Since then, however, there have been some spectacular examples of ADRs occurring with newly introduced drugs or formulations, including practolol, benoxaprofen, tienilic acid, zimelidine, Osmosin, and temafloxacin. More recently, there have been the major safety issues associated with anorexigens¹³ and third generation oral contraceptives.^{14,15} Those concerned with the release of new drugs, either in the pharmaceutical industry or in regulatory agencies, are now keen to have available monitoring systems that will detect such potentially serious ADRs quickly and efficiently.¹⁶ Unfortunately the intensive hospital-based monitoring system, as described above, is *not* such a scheme. It is expensive, is applicable only to drugs used frequently in hospitals, and, most importantly, the major data set extant of this type is unlikely to be useful in answering currently important questions. Other groups, however, have continued to collect and analyze data and, although there remains a fundamental problem with speed, the basic value of the method is repeatedly demonstrated. Details are given in a subsequent section.

By definition, intensive hospital based cohort studies are used to study drugs as they are used in a restricted (hospitalized) population. Thus, drugs that are primarily used on a long term basis in the community will only be studied indirectly when, for example, patients taking these preparations are

admitted to a hospital. The concept of a centralized collaborative approach is, however, valuable when dealing with a specific drug that has significant safety concerns, whether it is used in primary or secondary care. One such example is clozapine. Local monitoring centers can link with national information resources and can also feed back to prescribers and health care providers, thus improving the effectiveness of therapy and patient care.¹⁷ This is a general application of the target drug approach as detailed below. In addition, even for drugs that are used in hospital, a new drug will usually take some time to penetrate the market (depending, of course, on advertising and sales promotion). Thus, a relatively infrequent ADR is likely to go undetected for long periods of time within a small, restricted hospital based program, particularly if the ADR is delayed in onset. Such an ADR might be occurring at an unacceptably high rate. For example, aplastic anemia occurring with a frequency of one case per 5000 courses of treatment would be considered unacceptable for a new drug, yet such a frequency could not be detected using this method. Routine review of the data may yield an occasional interesting case, but it would certainly not be possible to establish a large enough body of evidence to suggest a causal link. This type of monitoring system is therefore not in direct competition with the spontaneous reporting schemes for ADRs operating in the US,¹⁸ in UK,¹⁹ and elsewhere (see Chapters 10 and 11), nor with the well established community based record linkage cohort studies (see Chapters 15–23).

This type of approach works well in the general hospital environment, where details of patient presentation, investigation, treatment, and prescribing are all relatively standardized and well recorded. In such settings, ward monitors can document all important and relevant information on events, prescribing, and possible ADRs without too much difficulty. For example, drug prescription orders tend to be changed no more than once per day and illnesses tend to follow a predictable course under these circumstances. This relatively standardized information can easily be checked, validated, and entered onto a computer file. It is also feasible and practical to extract this information and analyze it.

However, the system of ward monitoring as described is not well suited to all types of modern hospital practice. For example, any attempt to monitor patients in an intensive care unit by these techniques is unlikely to yield comprehensible data. Drug orders may change by the hour and, unless a monitor is constantly present, may go unrecorded. Furthermore, the data accumulated in such cases would be extremely complex and difficult to analyze. Attempting to attribute events in such a situation to a particular drug or group of drugs would be fraught with difficulty and may be impossible. Therefore, either one would be left to collect and record all "events," whether drug attributed or not, or unacceptable errors or biases could creep into judgments regarding causality. Either way, subsequent analyses would be difficult, if not impossible, to undertake with any degree of confidence. Having said all that, computerized hospital systems for recording drug orders and administration do bring this a step closer. The ADE Prevention Study Group has monitored patients in medical and surgical units including intensive care units.²⁰ Similarly, hospital based reporting systems have been used to study putative adverse drug effects, e.g., those related to cardiopulmonary bypass.²¹ Thirteen percent of all cardiovascular adverse events in these patients were thought to be related to the administration of protamine, but only 19% of those were reported.

Intensive hospital based cohort studies require medical and epidemiologic expertise, trained monitors at all participating centers, data managers, data entry personnel, computer scientists, and secretarial help. In addition, many people are required to collect data. Such a program is therefore expensive to run. However, this cost must be viewed in light of the knowledge that can and has been gained, and in comparison with the cost of introducing a new chemical entity to the marketplace. Attempts have been made to quantify the costs attributed to ADRs.²² This is clearly an issue of importance to public health and to the pharmaceutical industry, as evidenced by the stringent safety requirements imposed by regulatory authorities.

Finally, and most importantly, BCDSP hospital cohort data are now approximately 15–20 years

old and provide us with no information on several new classes of drugs, such as H₂ receptor antagonists, calcium channel blockers, and angiotensin converting enzyme inhibitors. Moreover, many new individual drugs in previously studied classes have appeared, such as antibiotics and β -blockers, which have not been monitored by this approach. Thus, to be of continuing value, this approach needs a continuous data collection program. In recent years a number of smaller systems have provided useful information and continue to do so, including local experience with modern drugs.^{23,24}

PARTICULAR APPLICATIONS

BCDSP: STUDIES OF THE IN-HOSPITAL DRUG USE OF MEDICAL INPATIENTS

Drug Utilization Studies

Although studies of prescribing habits and drug utilization in hospitals were not the primary aims of the BCDSP, during routine analyses substantial differences in drug use patterns were noted among and within the countries participating in the program. One study was undertaken in which inpatient drug use in matched Scottish and American patients was compared and contrasted.²⁵ The American patients received twice as many drugs as did the Scots, and overall ADR rates were correspondingly greater. This was despite the fact that the ADR rate for individual drugs was similar in the two countries.

Similarly, the use of intravenous fluid therapy was seen to vary widely among participating hospitals. Whereas 53% of patients admitted to one hospital in the US received intravenous fluids, only 7% of patients admitted to an Israeli hospital did so. Even within a country with low overall drug usage such as Scotland, two medical units within one city with otherwise virtually identical drug use patterns showed a twofold difference in intravenous fluid use.²⁶ Studies of this type highlight the need for careful studies of the reasons for prescribing expensive drugs if overall drug costs are to be held in check. A comprehensive data set

is essential if meaningful conclusions are to be made regarding differences in use and expenditure patterns between institutions, departments, or prescribers.

Information from hospital based cohort studies on the most commonly prescribed drugs, their indications for use, and the most common ADRs experienced by patients receiving them have been published in detail elsewhere.²⁷

Descriptive Studies of Adverse Drug Reactions

Drug related deaths are recorded infrequently in medical inpatients. In a major review of patients studied during the early years of the program, only 24 drug attributed deaths were found among 24 462 consecutive admissions.²⁸ Most of these were very ill patients in whom death was the expected outcome of hospitalization. In only six cases did it seem possible that the death could have been prevented. These were five patients who experienced fluid overload and one who had hyperkalemia from excessive potassium therapy.

With such large numbers of patients under study, it is possible to look for uncommon or rare drug effects that occur after short term drug use. For example, 119 cases of anaphylaxis, convulsions, deafness, or extra-pyramidal symptoms were found among 38 812 patients who received over 250 000 courses of drug treatment.²⁹ As might be expected, very few drugs were found to be responsible, but for those that were, it was possible to estimate incidence rates, albeit with wide confidence intervals.

By regularly reviewing the accumulating information, it was possible to identify subgroups of the population who are at a greater than expected risk of developing ADRs. Thus, for example, in an early study of heparin, women, especially those over 60 years old, were found to be at greatest risk of bleeding.³⁰ The strength of the association was such that the observation was made after only 97 patients had received the drug, further recipients serving to confirm the association. More recently the original observations have been confirmed and extended, by reviewing data on 2656 patients receiving heparin therapy. Bleeding was a dose related phenomenon occurring most often in

women, severely ill patients, and patients receiving aspirin during heparin therapy. The 7 day cumulative risk for bleeding was 9.1%.³¹

Similarly flurazepam, a commonly used hypnotic, was found to have a modest overall ADR rate of 3.1%, the most common ADR being excessive drowsiness or "hangover" on the morning following use.³² However, when the results were reviewed according to age and dose, it was clearly demonstrated that this drug had a very high ADR rate in the elderly, especially when high doses were used.

One particularly elegant study revealed a large number of factors associated with acute pharmacological effects of the hypotensive agent methyldopa.³³ In everyday use in the series of university teaching hospitals covered by the BCDSP, an average of 10% of 1067 methyldopa recipients experience clinically significant degrees of hypotension on starting treatment. Detailed analyses of these subjects showed a strong age relationship: the younger the patient, the greater the frequency of hypotension. There were also independent positive associations between hypotension and measurements of patient weight, daily drug dose, degree of hypertension on admission, and kidney function. The magnitude of these associations was considerable. Thus, for example, there was a sixfold greater frequency of drug associated hypotension in patients with renal impairment receiving a high initial dose of methyldopa, when compared with patients without renal impairment receiving low doses of this drug. These associations allowed the group to make recommendations for altering dosing regimens when starting treatment with methyldopa. There are several important points that should be noted about this study. First, the results were biologically and pharmacologically plausible. Second, the effects were predictable pharmacologic ones. Third, they were occurring during everyday use of a common medication. Finally, they were preventable with careful alteration in dosing.

Other studies of considerable clinical interest and relevance included those on potassium chloride,³⁴ furosemide,³⁵ spironolactone,³⁶ and the interrelationships between diuretic use and electrolyte responses in a large cohort of patients with congestive cardiac failure.³⁷

Hypothesis Testing Studies

On routine review of the accumulated data, certain relationships between ADRs and biochemical data were observed. For example, the rate of ADRs attributed to phenytoin was noted to be related to serum albumin concentration, the rate being 11.4% when serum albumin was less than 30 g l^{-1} and 3.8% when serum albumin was greater than 30 g l^{-1} .³⁸ This relationship was independent of age, dose of drug, and renal function, and was likely to be due to phenytoin being both strongly protein bound and of a low therapeutic index. Thus, a small decrease in binding protein may lead to a large increase in free phenytoin concentrations and toxic effects.

Although drug interactions are not often a major clinical issue, they can be found in data of this type. For example, in 1970 it was suggested that the hypnotic chloral hydrate might enhance the anticoagulant activity of warfarin, the proposed mechanism being the displacement of warfarin from binding sites by the metabolite trichloroacetic acid. The BCDSP data were ideal for testing this hypothesis, as they had been gathered before the hypothesis was put forward and prescribing doctors were therefore unaware of a potential interaction, that is there was no prescribing bias. To test this hypothesis, warfarin recipients were divided into three groups depending upon whether they received regular, intermittent, or no chloral hydrate during admission. The amount of warfarin required to maintain adequate anticoagulation was found to be inversely related to chloral hydrate ingestion, that is those taking chloral hydrate regularly required less warfarin. Thus, the original hypothesis was confirmed.³⁹

Similarly, isolated case reports on the interaction between phenytoin and the antituberculous drug isoniazid led to concern about their co-administration. Review of the BCDSP data revealed that six of 22 patients receiving these two drugs concurrently experienced central nervous system adverse effects (27%). By contrast, only 30 of 1093 phenytoin recipients who did not receive isoniazid experienced these adverse effects

(3%). Thus, these data confirmed the initial suggestion from the (uncontrolled) spontaneous reports.⁴⁰

In an interesting and important study, Duhme and his colleagues⁴¹ compared the adverse reaction rates attributed to digoxin in two Boston hospitals. Reactions were reported in 10% of 272 patients at one hospital and in only 4% of 291 patients at another. The differences did not appear to be due to different digoxin dosage, use of other drugs, or measured patient characteristics. They also did not appear to be related to reporting bias. In the hospital with the lower frequency of digoxin toxicity, serum digoxin levels were monitored more often (40% of patients) than in the other hospital (12% of patients). This led the investigators to conclude that frequent use of serum digoxin assays in clinical practice could decrease the frequency with which recipients experience adverse reactions.

In 1970 it was suggested that hospital patients with cardiac disease who were treated with tricyclic antidepressant drugs suffered excess mortality. Detailed review of the BCDSP data on hospitalized medical patients showed a high overall rate of ADRs in tricyclic antidepressant recipients, but also showed that the rate of sudden death in tricyclic recipients (0.4%) was similar to that of nonrecipients (0.3%).⁴²

Smoking might be expected to influence drug metabolism since substances in cigarette smoke can induce hepatic microsomal activity. It was also possible to test this hypothesis in a number of ways. The effect of the analgesic drug propoxyphene was found to be inversely related to the amount of cigarette smoking, regardless of the condition that required analgesic therapy.⁴³ Similarly, among users of diazepam and chlordiazepoxide, drug attributed drowsiness was less common in smokers than nonsmokers. This effect was not seen in phenobarbital users. It was suggested that users of phenobarbital already had maximally induced enzymes, and that smoking therefore made no difference.⁴⁴

Other hypotheses tested included relationships between tetracycline use and renal impairment (confirmed),⁴⁵ increased nephrotoxicity in patients receiving both aminoglycoside and cephalosporin

(not confirmed),⁴⁶ and amphetamine use and Hodgkin's disease (not confirmed).⁴⁷

Hypothesis Generating Studies

Regular review of the BCDSP data often produced observations which led to hypotheses about previously unsuspected drug reactions. For example, after a short period of monitoring, when only 4000 patients had been enrolled, it was observed that gastrointestinal bleeding was more common in patients receiving intravenous ethacrynic acid.⁴⁸ The risk was present both in patients receiving heparin and in those who did not receive anticoagulants. Although other loop diuretics have gastrointestinal side effects, few, if any, are associated with gastrointestinal bleeding.

An interesting example of the detection of increased susceptibility to ADRs was observed after the BCDSP was expanded to include monitoring centers in other countries. Skin rashes were found to be more common in Israeli patients, especially those receiving the analgesic drug dipyrone.⁴⁹ The rates were 2.3% for American patients, 2.6% for Israeli patients not taking dipyrone, and 6.6% for Israeli patients taking dipyrone. This drug was given to a large number of Israeli patients and the rash often closely followed the start of drug therapy. The association was discovered from within the database, since the absolute frequency was low and the medical attendants rarely noticed the relationship.

In 1976, the first in a series of new analyses was undertaken, in which the drug incriminated by the attending physicians as the cause of a reaction was ignored. All 507 patients with a drug attributed skin rash among 22 227 consecutive patients were systematically reviewed. Having removed all information on recipients of penicillin and blood/blood products from the data set, and having defined certain rules of causality, 57 separate drugs were identified as having strong likelihood of causing skin reactions.⁵⁰

In a similar way, 57 cases of gastrointestinal hemorrhage in whom an underlying medical predisposition to bleed was unlikely and a drug related cause seemed possible were identified among 16 646 medical inpatients.⁵¹ This review

indicated a strong likelihood that steroids, aspirin, anticoagulants, and ethacrynic acid could cause gastrointestinal hemorrhage in otherwise healthy subjects. It must be remembered that, although this seems a small number of cases, modern nonsteroidal anti-inflammatory drugs were not in use at the time.

BCDSP: STUDIES OF THE PREHOSPITAL DRUG USE OF MEDICAL INPATIENTS

A review of the proportion of patients admitted to medical wards due to ADRs was published in 1974.⁵² This emphasized that between 1.8 and 5.6% of admissions to the medical wards of seven university teaching hospitals were due to such reactions. Foremost among the drugs associated with this type of reaction were digoxin, aspirin containing preparations, insulin, anticoagulants, and adrenal corticosteroids. However, the period of interest was initially restricted to the 3 months preceding hospitalization, as it was felt that patient recall of drug history for a longer period would be inaccurate. Regular drug use which stopped for a period exceeding three months before admission was not recorded, and any relationship between this drug use and subsequent morbid events was undetected. Modern approaches to this method often have access to computerized prehospital drug use information (see Chapters 15–23), thus eliminating potential recall bias.

This limitation may have been relevant to a study that sought an association between prehospitalization aspirin use and renal disease.⁵³ No such association was found. However, the exposure status in some patients with renal disease may have been distorted by the 3 month rule. Such patients could have consulted a physician many months previously because of symptoms from renal disease and been advised to discontinue aspirin. If they complied with this request and then an admission to the hospital occurred later, a history of exposure to analgesics could have gone unrecorded. This problem could have been circumvented by confining analyses to newly diagnosed cases of the condition under review.

During the study of prehospital aspirin consumption, it was observed that there was a

significantly lower mortality than expected in this group of patients. Further analyses revealed that this was due to a reduction in myocardial infarction.⁵⁴ This finding aroused considerable interest having, as it did, major implications for prevention. The work was repeated two years later, using a different group of monitored patients, with the same negative relationship being demonstrated.⁵⁵ Analyses took account of age, sex, hospital, and reason for drug therapy. Furthermore, the association was found in patients with a wide range of clinical conditions, such as diabetes, hypertension, previous myocardial infarction, and arthritis. Details of potential risk factors such as diet, exercise habits, and lifestyle were not recorded. Although these could represent confounding factors, none were felt likely to significantly alter the result. Similarly, as only myocardial infarction patients who survived to reach a hospital could be studied, a theoretical explanation might be that aspirin takers were more likely to die in the acute phase of infarction. This seemed unlikely to explain the findings, however. It was to be over ten years before this finding was confirmed in a subsequent randomized clinical trial, and of course, aspirin is now a very important therapeutic tool in prevention and treatment of cardiovascular disease.

There exists within the framework of such an elaborate system for data collection the possibility of gathering information on substances which, although not drugs or medications in the accepted sense, may nevertheless have pharmacological actions or interactions with other drugs. Information on tea, coffee, and alcohol consumption was collected by the group. From the regular review of this information came the observation that regular coffee drinking was associated with nonfatal myocardial infarction.⁵⁶ Risk was found to be increased twofold and was found to be independent of age, sex, previous myocardial infarction, other cardiac conditions, obesity, diabetes, smoking, and occupation. Further corroborative evidence was obtained in a separate cross-sectional study carried out in 1972 (see below), where risks of the same magnitude were found.⁵⁷ Various explanations for the finding are theoretically possible.

Using a similar (case-control) approach, an association was sought between alcohol consumption and myocardial infarction, but neither an increase nor a decrease in risk due to alcohol consumption was demonstrated.⁵⁸

Starting at a time when quantitative information about ADRs was poor or nonexistent, the BCDSPP has been the most successful group to study hospital based cohorts of drug recipients and to quantify the adverse effects of established drugs. A summary of the studies reported by these workers was published in 1986.⁵⁹ By 1974, its Director, Hershel Jick, was able to review his work at BCDSPP and conclude that, although ADRs are common and may affect many millions of Americans annually, their severity and the rates with individual drugs are remarkably low.⁶⁰ After that time, the emphasis of research work at BCDSPP shifted to different information sources (see Chapters 15 and 23), although *ad hoc* studies are still conducted from time to time using their original methodology.

Other Intensive Hospital Based Cohort Studies

Although the largest proportion of analyses from the BCDSPP involved medical inpatients, detailed studies have been undertaken on smaller numbers of patients in pediatric wards¹⁰ and psychiatric wards.¹¹ Further analyses of the psychiatric resource revealed a strong negative association between chlorpromazine associated drowsiness and cigarette smoking habits.⁶¹ Thereafter, a modest monitoring program was undertaken on 5232 surgical patients.⁹ This study emphasized the higher number of drugs used in surgical patients when compared to medical patients and also indicated a measurable proportion of ADRs occurring in these patients. Further detailed reviews of this surgical data resource defined the prevalence of acute respiratory events occurring after general anesthesia in the recovery room.⁶²

In addition, although the BCDSPP has accumulated the largest database using short term in-hospital cohort studies, it is by no means the only group that has been using this approach. The initial pediatric drug monitoring studies

undertaken by Jick and his colleagues¹⁰ were soon discontinued, in large part because the information suggested that the rates of reactions were low in all but seriously ill young children. It was therefore concluded that the cost-effectiveness of this type of monitoring in pediatric wards was relatively low. Mitchell and his colleagues undertook a monitoring project in a large pediatric ward based on techniques similar to the original BCDSP study.⁶³ Several important findings emerged in addition to essential basic quantitation of drug utilization problems. In 1985, they published an important study showing a significant inverse relationship between body weight and risks of hyperglycemia following infusion with 10% dextrose solutions.⁶⁴ This emphasized the importance of careful dosage adjustment of treatments in the neonatal period and the necessity of continued vigilance even with apparently simple treatment regimens to avoid serious adverse effects. More recently, a review of the risks of heparin therapy used to monitor patency of venous access demonstrated a positive association with intraventricular hemorrhage in low birth weight infants.⁶⁵ In view of their findings, it is perhaps a cause for concern that these workers showed a significant increase in drug use in low birth weight infants when comparing data from 1978–79 with 1985–86, especially in the first week of treatment in the intensive care unit. In the latter phase, some 71% of admissions to the neonatal intensive care unit had received gentamicin during their stay, and approximately 60% each received sodium chloride, potassium chloride, and heparin. Such reports emphasize the need for continued careful monitoring of this vulnerable age group of patients. Recognizing the limitations of inpatient monitoring, the group has recently stressed the importance of the randomized controlled trial in monitoring for drug safety in children (see Chapter 33). In a large study, no increase in risk of hospitalization for gastrointestinal bleeding, renal failure, or anaphylaxis was found in short term users of ibuprofen.⁶⁷

In the field of general medical inpatient monitoring studies, those from the University of Berne, Switzerland, deserve special mention. In 1974, Hoigne and colleagues began routine systematic followup of all patients hospitalized in several

Berne medical units.⁶⁸ This program, known as the Comprehensive Hospital Drug Monitoring Berne (CHDMB), has published many interesting and valuable papers on ADRs.⁶⁹ Because of its size, this program is one of the few data resources whose work can be compared with that of the BCDSP. Their initial reports on drug related deaths in 17 285 medical patients in two teaching hospitals indicated an incidence of 0.4 per 1000 patients, a figure similar in magnitude to the 0.9 per 1000 patients recorded in a BCDSP study of 26 462 subjects.²⁸

CHDMB has published reports on drug induced blood dyscrasias⁷⁰ and on allergic reactions to drugs.⁷¹ CHDMB published two reports on skin reactions attributed to penicillins and their relationship to allopurinol use.^{72,73} These were of great methodologic interest because, although they confirmed the high frequency of exanthemata associated with penicillins reported by BCDSP,⁷⁴ they did not reveal an excess in patients concomitantly receiving penicillins and allopurinol, a finding initially coming from BCDSP.⁷⁵ The CHDMB reviewed all the published information and concluded that the most likely explanation for the discrepancy was variations in the length of exposure to penicillins in the two studies. An alternative explanation, however, could be that the CHDMB data set was a record of all penicillin or allopurinol attributed rashes, whereas the BCDSP data set was collected and analyzed to look at virtually all rashes occurring in the monitored wards. The judgments inherent in deciding whether to include a patient in the Berne data could have introduced unwanted bias into an otherwise excellent system. These findings highlight the need both for extreme care in the analyses of this type of information and clarity in the documentation of the final results.

Data collection continues in two centers in Berne and St Gallen. By 1993, the database contained information on 48 005 consecutive admissions of 34 840 patients. By recording prescription information from the 21 day prehospitalization period, the Swiss workers have contributed to the growing body of published information on the association between upper gastrointestinal bleeding and non-steroidal anti-inflammatory drugs.⁷⁶ Twenty years

of experience have been reviewed in recent reports of drug induced asthma and skin reactions.^{77,78} Older information from these databases has recently revealed further information on heparin induced thrombocytopenia, antibiotics/colitis, and diuretics/hypokalemia.^{79,80,81}

The general methodology found recent application in an Italian project running under the acronym ARIES (Adverse Reaction Identification Evaluation System). The project was developed as a collaborative effort between ICI-Pharma and several hospital departments that comprise the Gruppo di Ricerca sulla Epidemiologia del Farmaco. Physicians worked with external monitors in the collection of data of three forms: (i) demographic and administrative information together with ICD coded diagnoses, (ii) prescription information including drug name, route, dose, and starting and stopping dates, and (iii) adverse event information obtained using a simple algorithm. Data collected between 1988 and 1991 yielded information on 9000 patients and 60 000 prescriptions. Much of this early effort was directed towards drug utilization studies,⁸² but there existed within the framework the possibility of conducting hypothesis generating and testing studies similar to those conducted using the Boston and Swiss data.

THE FUTURE

Having highlighted some of the strengths and drawbacks of this approach, and provided some examples, it is relevant to conclude this review by looking ahead and attempting to assess how this technique of intensive hospital based monitoring might continue to contribute to our knowledge about ADRs.

First, patterns of drug use change as our understanding of pathophysiology advances, new indications for old drugs emerge, often after many years of research, and established drugs are used in new formulations. Furthermore, patterns of hospitalization change as economic and political developments proceed. Populations change as birth rates change and longevity increases. For all of these reasons, it is predictable that patterns of ADRs will change with time. It might therefore

be both reasonable and justifiable to conduct this type of intensive monitoring exercise every ten years or so, in order to keep abreast of current drug therapy and modes of clinical practice. The period over which monitoring might be done would not be crucially important. Indeed if a large number of centers were recruited and enough staff employed to cope with data handling, the required number of patients, say 50 000, could be processed quite quickly. Such a development would be both useful and justified. It would offer information of considerable value to prescribers, manufacturers, and regulators alike. However, the main advantage of periodically repeating the BCDSF exercise would be that all recently marketed drugs that have reached significant usage levels in hospitals could be monitored, their acute toxicity profile determined, and their interaction with other drugs and with smoking and alcohol habits assessed.

Realistically, cost is likely to be the main limiting factor. As hospital information systems improve, so does the extent and quality of information on drug prescribing and administration. Additional linkage to demographic, laboratory, procedures, and morbidity information raises the possibility of comprehensive multipurpose in-hospital information that would allow ongoing safety monitoring of the intensive in-hospital type, and indeed go a long way towards achieving the proposals for a scientifically rigorous system proposed by Finney in 1965.³

Were a general monitoring program to be instituted on a regular basis every decade or so, even then information on newly marketed drugs would remain scarce. However, this could be circumvented by selected (targeted) drug monitoring studies. Such studies have been undertaken many times in the past, and clozapine has already been cited as a more recent example.¹⁷ Target drug surveillance is a modification of the basic system whereby, rather than selecting a hospital service and scrutinizing all patients and all drugs in that service, a drug of interest is selected and all users of that drug, regardless of location within the hospital, are studied. For example, Koch-Weser and his colleagues at the Massachusetts General Hospital performed a classic study of this type, reviewing information on 317 patients receiving

colistin treatment. Detailed analyses showed that altering the method of dose adjustment in patients with renal impairment would be likely to reduce the ADR rate significantly.⁸³ This system has been used by the Boston group to study various drugs, including intravenous cimetidine⁸⁴ and atracurium, a neuromuscular blocking agent for use in anesthesia.^{85,86} The work of the monitor in such a system is clearly different, for there must be close liaison with the hospital pharmacy if all drug users are to be identified and located. In this situation the monitor works throughout the hospital and interacts with many more people than his or her counterpart in the previously described system, so such a monitor must be both highly efficient and a good communicator.

The target drug system is certainly a more efficient way of answering specific questions about new drugs, and is therefore a more efficient approach for this defined purpose than the general inpatient monitoring system. It also lends itself to the multicenter approach. The Drug Surveillance Network is a nationwide network of clinical pharmacists in the USA and Canada. Clinical pharmacists in over 300 hospitals record relevant information and monitor progress and outcomes in consecutive eligible patients. Studies on antibiotic use have highlighted potential problems with renal and hematological ADRs, and methodological issues related to misclassification have also been addressed.^{87,88}

Another way of focusing resources in an intensive hospital based monitoring program is by selecting specific patient groups who, for various reasons, may be worthy of particularly detailed study. The pediatric monitoring program conducted by Mitchell was referred to earlier.⁶³⁻⁶⁶ An additional example of such a group would be elderly persons, who are at considerable risk of developing ADRs by virtue of high drug usage, if for no other reason. Subjects over the age of 65 years comprise about 15% of the UK population and consume about 35% of prescribed drugs. They also comprise over half the total hospitalized population. ADRs in the elderly are common. Hurwitz⁸⁹ provided documentation of this in her study of medical inpatients in Belfast. Williamson and Chopin⁹⁰ studied a group of admissions to

geriatric assessment units and predictably found a high rate of ADRs. They documented the drugs most frequently responsible and demonstrated a link between prehospitalization polypharmacy and ADRs. A significant proportion of hospital admissions is thought, at least in part, to be due to ADRs.⁹¹ There remains overall concern that elderly patients are at considerable risk from ADR,^{92,93} although the complex relationship between age, altered physiology, multiple drug exposures, and multiple disease states needs careful interpretation.

Monitoring systems for assessing ADRs in elderly subjects have largely been confined to acute medical wards, although community based record linkage gives exciting opportunities to investigate drug exposure and outcomes in a very different group. Few studies have investigated elderly subjects in chronic care institutions, a group at high risk from serious reactions.⁹⁴ A program gathering such data on a routine basis would be desirable, and the intensive hospital based system could be adapted for use in this environment.

In a detailed review of the "state of the art" of the discovery of drug induced illness, the Director of the BCDSF defined the contribution of various different research strategies to this end.⁹⁵ He emphasized that the cohort approach was most useful in situations where a drug causes a significant increase in an already high baseline risk of an event or illness. Such situations are relatively unusual, albeit still of major importance. This view was upheld by Venning⁹⁶⁻¹⁰⁰ who undertook a detailed study of identification of adverse reactions to new drugs.

Finally, in the US the Joint Commission for Accreditation of Hospitals now requires all hospitals to develop the capability to document and evaluate ADRs that occur in their patients. This could simply involve the passive collection of spontaneous reports of adverse reactions occurring in the hospitals. Alternatively, it could result in systems as thorough as the existing systems for monitoring for nosocomial infections or even as thorough as the intensive hospital based cohort studies described here.

It seems likely therefore that intensive hospital based cohort studies of suspected adverse drug

reactions and of events in drug recipients will remain a relevant technique for studying selected drugs in specific circumstances for some time. General monitoring of medical inpatients by this approach would be useful if undertaken periodically by a trained research team, and may be a useful by-product of rapidly developing information systems.

REFERENCES

- Best WR. Chloramphenicol-associated blood dyscrasias. A review of cases submitted to the American Medical Association Registry. *J Am Med Assoc* 1967; **201**: 181–8.
- Lenz W. Thalidomide and congenital abnormalities. *Lancet* 1962; **1**: 45.
- Finney DJ. The design and logic of a monitor of drug use. *J Chron Dis* 1965; **18**: 77–98.
- Seidl LG, Thornton GF, Smith JW, *et al.* Studies on the epidemiology of adverse drug reactions. III. Reactions in patients on a general medical service. *Bull Hopkins Hosp* 1966; **119**: 299–315.
- Thoburn R, Johnson JE, Cluff LE. Studies on the epidemiology of adverse drug reactions. IV. The relationship of cephalothin and penicillin allergy. *J Am Med Assoc* 1966; **198**: 345–8.
- Hurwitz N, Wade OL. Intensive hospital monitoring of adverse reactions to drugs. *Br Med J* 1969; **1**: 531–6.
- Gardner P, Watson LJ. Adverse drug reactions: A pharmacist-based monitoring system. *Clin Pharmacol Ther* 1970; **11**: 802–7.
- Slone D, Jick H, Borda I, *et al.* Drug surveillance utilizing nurse monitors. *Lancet* 1966; **2**: 901–3.
- Boston Collaborative Drug Surveillance Program. Drug monitoring of surgical patients. *J Am Med Assoc* 1982; **248**: 1482–5.
- Lawson DH, Shapiro S, Slone D, *et al.* Drug surveillance: problems and challenges. *Pediatr Clin N Am* 1972; **19**: 117–29.
- Swett C. Outpatient phenothiazine use and bone marrow depression. *Arch Gen Psychiatry* 1975; **32**: 1416–8.
- Jick H, Miettinen OS, Shapiro S, *et al.* Comprehensive drug surveillance. *J Am Med Assoc* 1970; **213**: 1455–60.
- Connolly HM, Crary JL, McGoon MD, Hensrud DD, Edwards BS, Edwards WDS, Schaff HV. Valvular heart disease associated with fenfluramine–phentermine. *New Engl J Med* 1997; **337**: 581–8.
- Spitzer WO, Lewis MA, Heinemann LAJ, Thorogood M, MacRae KD. Third generation oral contraceptives and risk of venous thromboembolic disorders: an international case–control study. *Br Med J* 1996; **312**: 83–8.
- Jick H, Jick SS, Gurewich V, Myers MW, Vasilakis C. Risk of idiopathic cardiovascular death and nonfatal venous thromboembolism in women using oral contraceptives with differing progestagen components. *Lancet* 1995; **346**: 1589–93.
- Grahame-Smith DG. *Report of the Adverse Reaction Working Party to the Committee on Safety of Medicines*. London: Department of Health and Social Security, 1986.
- Black LL, Greenidge LL, Ehmann T, Ganesan S, Honer WG. A centralized system for monitoring clozapine use in British Columbia. *Psychiatr Serv* 1996; **47**: 81–3.
- Rossi AC, Hsu JP, Faich GA. Ulcerogenicity of piroxicam: an analysis of spontaneously reported data. *Br Med J* 1987; **294**: 147–50.
- Rawlins MD. Spontaneous reporting of adverse drug reactions. *Q J Med* 1986; **230**: 531–4.
- Cullen DJ, Sweitzer BJ, Bates DWB, E, Edmondson A, Leape LL. Preventable adverse drug events in hospitalized patients: a comparative study of intensive care and general care units. *Crit Care Med* 1997; **25**: 1289–97.
- Kimmel SE, Sekeres MA, Berlin JA, Goldberg LR, Strom BL. Adverse events after protamine administration in patients undergoing cardiopulmonary bypass: risks and predictors of under-reporting. *J Clin Epidemiol* 1998; **51**: 1–10.
- Bates DW, Spell N, Cullen DJ, Burdick EL, N, Petersen LA, Small SD, *et al.* The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. *J Am Med Assoc* 1997; **277**: 307–11.
- Moore N, Lecoindre D, Noblet C, Mabille M. Frequency and cost of serious adverse drug reactions in a department of general medicine. *Br J Clin Pharmacol* 1998; **45**: 301–8.
- Smith CC, Bennett PM, Pearce HM, Harrison PI, Reynolds DJ, Aronson JKG, Grahame-Smith DG. Adverse drug reactions in a hospital general medical unit meriting notification to the Committee on Safety of Medicines. *Br J Clin Pharmacol* 1996; **42**: 423–9.
- Lawson DH, Jick H. Drug prescribing in hospitals: an international comparison. *Am J Public Health* 1977; **66**: 644–8.
- Lawson DH. Intravenous fluids in medical inpatients. *Br J Clin Pharmacol* 1977; **4**: 299–303.
- Miller RR. Drug surveillance utilizing epidemiologic methods. *Am J Hosp Pharm* 1973; **30**: 584–92.

28. Porter J, Jick H. Drug related deaths among medical inpatients. *J Am Med Assoc* 1977; **237**: 279–81.
29. Porter J, Jick H. Drug-induced anaphylaxis, convulsions, deafness and extrapyramidal symptoms. *Lancet* 1977; **1**: 587–8.
30. Jick H, Slone D, Borda I, *et al.* Efficacy and toxicity of heparin in relation to age and sex. *New Engl J Med* 1968; **279**: 284–6.
31. Walker AM, Jick H. Predictors of bleeding during heparin therapy. *J Am Med Assoc* 1980; **244**: 1209–12.
32. Greenblatt DJ, Allan MD, Shader RI. Toxicity of high dose flurazepam in the elderly. *Clin Pharmacol Ther* 1977; **21**: 355–61.
33. Lawson DH, Gloss D, Jick H. Adverse reactions to methyl dopa with special reference to hypotension. *Am Heart J* 1978; **96**: 572–9.
34. Lawson DH. Adverse reactions to potassium chloride. *Q J Med* 1974; **43**: 433–40.
35. Greenblatt DJ, Duhme DW, Allen MD, *et al.* Clinical toxicity of furosemide in hospitalized patients. *Am Heart J* 1977; **94**: 6–13.
36. Greenblatt DJ, Koch-Weser J. Adverse reactions to spironolactone. *J Am Med Assoc* 1973; **225**: 40–3.
37. Lawson DH, O'Connor PC, Jick H. Drug attributed alterations in potassium handling in congestive cardiac failure. *Eur J Clin Pharmacol* 1982; **23**: 21–5.
38. Boston Collaborative Drug Surveillance Program. Diphenylhydantoin side effects and serum albumin levels. *Clin Pharmacol Ther* 1973; **14**: 529–32.
39. Boston Collaborative Drug Surveillance Program. Interaction between chloral hydrate and warfarin. *New Engl J Med* 1972; **286**: 53–5.
40. Miller RR, Porter J, Greenblatt DJ. Clinical importance of the interaction of phenytoin and isoniazid. *Chest* 1979; **75**: 356–8.
41. Duhme DW, Greenblatt DJ, Koch-Weser J. Reduction of digoxin toxicity associated with measurement of serum levels. *Ann Intern Med* 1974; **80**: 516–9.
42. Boston Collaborative Drug Surveillance Program. Adverse reactions to the tricyclic antidepressant drugs. *Lancet* 1972; **1**: 529–31.
43. Boston Collaborative Drug Surveillance Program. Decreased clinical efficacy of propoxyphene in cigarette smokers. *Clin Pharmacol Ther* 1973; **14**: 259–63.
44. Boston Collaborative Drug Surveillance Program. Clinical depression of the central nervous system due to diazepam and chlordiazepoxide in relation to cigarette smoking and age. *New Engl J Med* 1973; **288**: 277–80.
45. Boston Collaborative Drug Surveillance Program. Tetracycline and drug-attributed rises in blood urea nitrogen. *J Am Med Assoc* 1972; **220**: 377–9.
46. Fanning WL, Gump D, Jick H. Gentamicin and cephalothin-associated rises in blood urea nitrogen. *Antimicrob Agents Chemother* 1976; **10**: 80–2.
47. Boston Collaborative Drug Surveillance Program. Amphetamines and malignant lymphoma. *J Am Med Assoc* 1974; **229**: 1462–3.
48. Slone D, Jick H, Lewis GP, *et al.* Intravenously given ethacrynic acid and gastrointestinal bleeding. *J Am Med Assoc* 1969; **209**: 1668–71.
49. Boston Collaborative Drug Surveillance Drug Program. Dipyron as a cause of drug rashes: an epidemiologic study. *Int J Epidemiol* 1973; **2**: 167–70.
50. Arndt KA, Jick H. Rates of cutaneous reactions to drugs. *J Am Med Assoc* 1976; **235**: 918–23.
51. Jick H, Porter J. Drug-induced gastrointestinal bleeding. *Lancet* 1978; **2**: 87–9.
52. Miller RR. Hospital admissions due to adverse drug reactions. *Arch Intern Med* 1974; **134**: 219–24.
53. Lawson DH. Analgesic consumption and impaired renal function. *J Chron Dis* 1973; **26**: 39–45.
54. Boston Collaborative Drug Surveillance Program. Regular aspirin intake and acute myocardial infarction. *Br Med J* 1974; **1**: 440–3.
55. Jick H, Miettinen OS. Regular aspirin use and myocardial infarction. *Br Med J* 1976; **1**: 1057–60.
56. Boston Collaborative Drug Surveillance Program. Coffee drinking and acute myocardial infarction. *Lancet* 1972; **2**: 1278–81.
57. Jick H, Miettinen OS, Neff RK, *et al.* Coffee and myocardial infarction. *New Engl J Med* 1973; **289**: 63–7.
58. Stason WB, Neff RK, Miettinen OS, *et al.* Alcohol consumption and nonfatal myocardial infarction. *Am J Epidemiol* 1976; **104**: 603–8.
59. Cohen MR. A compilation of abstracts and index published by the Boston Collaborative Drug Surveillance Program 1966–1985. *Hosp Pharm* 1986; **21**: 497–559.
60. Jick H. Drugs—remarkably nontoxic. *New Engl J Med* 1974; **291**: 824–8.
61. Swett C. Adverse reactions to chlorpromazine in psychiatric patients. *Dis Nerv Syst* 1974; **35**: 509–11.
62. Beard K, Jick H, Walker AM. Adverse respiratory events occurring in the recovery room after general anesthesia. *Anesthesiology* 1986; **64**: 269–72.
63. Mitchell AA, Lovejoy FH, Slone D, *et al.* Acetaminophen and aspirin. Prescription use and accidental ingestion among children. *Am J Dis Child* 1982; **136**: 976–9.
64. Louik C, Mitchell AA, Epstein MF, Shapiro S. Risk factors for neonatal hyperglycaemia associated with 10% dextrose infusion. *Am J Dis Child* 1985; **139**: 783–6.
65. Lesko SM, Mitchell AA, Epstein MF, Louik C, Giacoia SP, Shapiro S. Heparin use as a risk factor

- for intraventricular hemorrhage in low-birth weight infants. *New Engl J Med* 1986; **314**: 1156–60.
66. Lesko SM, Epstein MF, Mitchell AA. Recent patterns of drug use in newborn intensive care. *J Pediatr* 1990; **116**: 985–90.
67. Lesko SM, Mitchell AA. An assessment of the safety of pediatric ibuprofen. A practitioner-based randomized clinical trial. *J Am Med Assoc* 1995; **273**: 929–33.
68. Zoppi M, Torok M, Stoller-Guleryuz D, et al. Adverse drug reactions as probable causes of death. Results of the Comprehensive Hospital Drug Monitoring Berne (CHDMB). *Swiss Med J* 1982; **112**: 1808–10.
69. Torok M, Zoppi M, Winzenried P, et al. Drug related death among 17, 285 inpatients in the divisions of internal medicine of two teaching hospitals in Berne 1974–1980. In: Manell P, Johansson SG, eds, *The Impact of Computer Technology on Drug Information*. Amsterdam: North-Holland, 1982; 79–84.
70. Baumgartner A, Hoigne R, Muller U, et al. Drug induced blood dyscrasias. Results of the Comprehensive Hospital Drug Monitoring Berne (CHDMB) 1974–1979. *Swiss Med J* 1982; **112**: 1530–9.
71. Hoigne R, Stocker F, Middleton P. Epidemiology of drug allergy: drug monitoring. In: de Weck AL, Bundgaard H, eds, *Handbook of Experimental Pharmacology Vol. 63, Allergic Reactions to Drugs*. Berlin: Springer, 1983; 187–204.
72. Sonntag MR, Zoppi M, Fritschy D, et al. Exantheme unter haufig angewandten Antibiotika und antibakteriellen Chemotherapeutika (Penicilline, speziell Aminopenicilline, Cephalosporine und Cotrimoxazol) sowie Allopurinol. *Swiss Med J* 1986; **116**: 142–5.
73. Hoigne R, Sonntag MR, Zoppi M, et al. Occurrence of exanthems in relation to aminopenicillin preparations and allopurinol [letter]. *New Engl J Med* 1987; **316**: 1217.
74. Shapiro S, Slone D, Siskind V, et al. Drug rash with ampicillin and other penicillins. *Lancet* 1969; **2**: 969–72.
75. Boston Collaborative Drug Surveillance Program. Excess of ampicillin rashes associated with allopurinol or hyperuricemia. *New Engl J Med* 1972; **286**: 505–7.
76. Stodolnik E, Maurer P, Hoigne R, Hess T, Muller U, Amonn F, Halter F, Maibach R, Kunzi UP. Risk of acute upper gastrointestinal bleeding in patients with ulcerative disease and treatment with nonsteroidal anti-inflammatory drugs (NSAIDs). *Eur J Clin Pharmacol* 1990; **38**: 31–5.
77. Hoigne RV, Braunschweig S, Zehnder D, Kuenzi UP, Hess TL. Drug-induced attack of bronchial asthma in inpatients: a 20-year survey of the Comprehensive Hospital Drug Monitoring Programme on adverse drug reactions, Berne/St. Gallen. *Eur J Clin Pharmacol* 1997; **53**: 81–2.
78. Hunziker T, Kunzi UP, Braunschweig S, Zehnder D, Hoigne R. Comprehensive hospital drug monitoring (CHDM): adverse skin reactions, a 20-year survey. *Allergy* 1997; **52**: 388–93.
79. Killer S, Zoppi M, Kunzi UP, Hess T, Galeazzi RL, Muller UH, Hoigne R. Heparin-induced thrombocytopenia. Results from the Comprehensive Hospital Drug Monitoring (CHDM) Bern/St. Gallen. *Schweizerische Medizinische Wochenschrift* 1995; **125**: 2518–22.
80. Zehnder D, Kunzi UP, Maibach R, Zoppi M, Halter F, Neftel KA, et al. Frequency of antibiotics-associated colitis in hospitalized patients in 1974–1991 in Comprehensive Hospital Drug Monitoring, Bern/St. Gallen. *Schweizerische Medizinische Wochenschrift* 1995; **125**: 676–83.
81. Widmer P, Maibach R, Kunzi UP, Capaul R, Mueller U, Galeazzi RH, R. Diuretic-related hypokalaemia: the role of diuretics, potassium supplements, glucocorticoids and beta 2-adrenoceptor agonists. Results from the Comprehensive Hospital Drug Monitoring Programme, Berne (CHDM). *Eur J Clin Pharmacol* 1995; **49**: 31–6.
82. Maestrello I, Bissoli F, Croce G, Fassina S, Michela V, Giuliani R. La sorveglianza delle reazioni avverse da farmaci e il monitoraggio delle prescrizioni mediche come attivita “normali” di reparti ospedalieri. *Recenti Prog Med* 1991; **82**: 197–201.
83. Koch-Weser J, Sidel VW, Federman EB, et al. Adverse effects of sodium colistimethate. *Ann Intern Med* 1970; **72**: 857–68.
84. Porter JB, Beard K, Walker AM, et al. Intensive hospital monitoring study of intravenous cimetidine. *Arch Intern Med* 1986; **146**: 2237–9.
85. Jick H, Andrews EB, Tilson HH, et al. Atracurium—a post-marketing surveillance study. Part I. Methods and US experience. *Br J Anaesth* 1989; **62**: 590–5.
86. Lawson DH, Paice GM, Glavin RJ, et al. Atracurium—a post-marketing surveillance study. Part II. The UK study and discussion of findings. *Br J Anaesth* 1989; **62**: 596–600.
87. Timm EG, Welage LS, Walawander CA, Sayers JF, Karpiuk EL, Davis TDG Jr. Adverse drug reaction reporting in a multicenter surveillance study. *Ann Pharmacother* 1995; **29**: 240–5.
88. Welage LS, Walawander CA, Timm EG, Grasela TH. Risk factors for acute renal insufficiency in patients with suspected or documented bacterial pneumonia. *Ann Pharmacother* 1994; **28**: 515–22.
89. Hurwitz N. Predisposing factors in adverse reaction to drugs. *Br Med J* 1969; **1**: 536–9.

90. Williamson J, Chopin JM. Adverse reactions to prescribed drugs. *Age Ageing* 1980; **9**: 73–80.
91. Cunningham G, Dodd TR, Grant DJ, McMurdo ME, Richards RM. Drug-related problems in elderly patients admitted to Tayside hospitals, methods for prevention and subsequent reassessment. *Age Ageing* 1997; **26**: 375–82.
92. Atkin PA, Shenfield GM. Medication-related adverse reactions and the elderly: a literature review. *Adverse Drug React Toxicol Rev* 1995; **14**: 175–91.
93. Beard K. Adverse reactions as a cause of hospital admission in the aged. *Drugs Aging* 1992; **2**: 356–67.
94. Ray WA, Taylor JA, Meador KG, Lichtenstein MJ, Griffen MR, Fought R, Adams ML, Blazer DG. Reducing antidepressant drug use in nursing homes. A controlled trial of provider education. *Arch Intern Med* 1993; **153**: 713–21.
95. Jick H. The discovery of drug-induced illness. *New Engl J Med* 1977; **296**: 481–5.
96. Venning GR. Identification of adverse reactions to new drugs. I: What have been the important adverse reactions since thalidomide? *Br Med J* 1983; **286**: 199–202.
97. Venning GR. Identification of adverse reactions to new drugs. II—How were 18 important adverse reactions discovered and with what delays? *Br Med J* 1983; **286**: 289–92.
98. Venning GR. Identification of adverse reactions to new drugs. II (continued): How were 18 important adverse reactions discovered and with what delays? *Br Med J* 1983; **286**: 365–8.
99. Venning GR. Identification of adverse reactions to new drugs. III: Alerting processes and early warning systems. *Br Med J* 1983; **286**: 458–60.
100. Venning GR. Identification of adverse reactions to new drugs. IV—Verification of suspected adverse reactions. *Br Med J* 1983; **286**: 544–7.

13

Case–Control Surveillance

SAMUEL SHAPIRO

Slone Epidemiology Unit, Boston University School of Medicine, Brookline, MA, USA

Secondly, the systems that have emanated from Boston—the initial system of intensive hospital monitoring and the new system of relatively nonspecific case–control enquiry—have made, or promise to make, valuable contributions

Richard Doll, 1977¹

The experience of being disastrously wrong is salutary; no epidemiologist should be denied it, and not many are.

Dennis Slone, 1975 (with apologies to John Kenneth Galbraith).²

Over 20 years have passed since 1977, when Richard Doll discussed the promise of case–control surveillance (CCS), then only 18 months old and little more than a concept.¹ In the intervening years CCS has contributed to the public health, and it has had a major impact on directions of research in the epidemiological evaluation of drugs, and indeed, on epidemiology in general (see the appendix: “Case–control surveillance. Publications 1975–1998”). CCS has discovered previously unsuspected relationships between drug use and disease risk, some beneficial and some adverse; CCS has also documented drug safety, and it has developed a large database for the testing of hypotheses proposed by others.

CCS came about as the end result of almost a decade of experience in the application of

epidemiological principles to the evaluation of drug effects. That effort, which commenced in earnest in the mid-1960s,³ benefited from the conceptual formulations of Finney,⁴ who described a statistical model of drug surveillance (which he termed a “drug monitor”), and from the field studies of Cluff.⁵

Finney and Cluff were among the first to appreciate that the drug explosion which followed the second world war, coupled with the occurrence of unanticipated adverse effects (e.g., phocomelia due to thalidomide),⁶ necessitated the development of surveillance systems. Finney proposed that drug surveillance should have the capacity to fulfill three functions: (i) to detect previously unsuspected drug/event associations (hypothesis generation); (ii) to validate such associations by in-depth analysis; and (iii) to test hypotheses proposed by others (hypothesis testing).

The foresight of Finney and Cluff was soon vindicated as a rapid succession of adverse effects were reported. Examples include oral contraceptives and venous thromboembolism,^{7,8} practolol and corneal perforation,⁹ *in utero* exposure to diethylstilbestrol and vaginal cancer,¹⁰ and estrogens and endometrial cancer.¹¹ Case–control methods played a critical role in the quantitative elucidation of most of these and many other

associations, and it was clear that for the foreseeable future they would continue in that role, particularly in relation to serious drug-induced morbidity.¹

The early successes of case-control methods were to influence the development of CCS, but at the time that Finney first proposed his model, their value as applied to drug safety was not yet widely appreciated. The classical studies of the oral contraceptive/thrombosis relationship,^{7,12} among others, helped to change that perception.¹³ Even then, however, it was still thought that the case-control approach was applicable only to *ad hoc* studies of single outcomes, that it lacked the capacity to discover unsuspected connections between drug use and multiple outcomes, and that it therefore had little to offer to the development of a drug surveillance system. It was thought that such a system must necessarily be based on followup methods. In addition, the general prejudice of the time still decreed that “prospective studies” were superior to “retrospective studies”—if, indeed, the latter had any value at all.¹⁴ Thus, the early attempts to organize surveillance systems explored followup methods, implemented mainly^{3,5,15} (although not exclusively)¹⁶ within hospital wards (see Chapter 12). Nurses were trained to enroll consecutive patients, to record all drug exposures that occurred during the hospital stay, to record confounding factors, to follow the patients, and to record events.

The first such system to operate successfully on a large scale was the Boston Collaborative Drug Surveillance Program (BCDSP).¹⁷ It was also the first to show that the three requirements of Finney’s drug surveillance model, or monitor, could be fulfilled. Computers were used to compare multiple drug exposures in relation to multiple events. Previously unsuspected associations could be identified, the associations could then be evaluated in depth, and hypotheses proposed by others could be assessed (see Chapter 12).

The demonstration that Finney’s drug surveillance monitor was feasible was an important accomplishment. With the passage of time, however, it became increasingly clear that hospital-based followup had the major limitation that it did not monitor the occurrence of serious diseases

attributable to drug use in the population at large. The system could not, for example, have documented any of the associations mentioned above. To correct this situation, a case-control component was added to the BCDSP methods. On enrollment, the patients were followed, as before, but now they were also interviewed for histories of regular drug use during the month before admission. Then, with the experience already gained, it was an easy step to carry out multiple case-control comparisons: each primary diagnosis that led to admission (cases) could be compared with all other admissions (controls) for the prior use of all drugs to which the patients had been exposed. Contrary to the accepted wisdom at that time, as a conceptual matter it seemed possible, after all, to use case-control methodology to assess multiple drug exposures in relation to multiple outcomes.

As expected, noncausal associations soon became apparent (e.g., insulin and diabetes). Then, as large numbers accumulated, it became possible to screen for unsuspected associations. One association, observed after some 9000 patients had been studied, was a significantly reduced rate of antecedent aspirin use among patients with myocardial infarction (cases), relative to all other diagnoses (controls). On more detailed analysis, and in particular after the exclusion from the control series of patients whose primary diagnoses were not independent of aspirin use (e.g., arthritis), the inverse association held.¹⁸ The finding was of considerable interest because it had been shown that aspirin inhibited platelet adhesiveness.¹⁹

To confirm the finding, and to evaluate risks of other diseases, a second study (colloquially, “the special study”) was soon organized. Nurses administered a standard questionnaire covering regular drug use in the three months before admission. They also recorded demographic data, and specific information on a few potentially confounding factors relevant to the aspirin/myocardial infarction association. In ten months they interviewed some 25 000 patients admitted to 24 hospitals, whose primary diagnoses were subsequently determined from the discharge summaries.

This study design had certain attractive features. First, because heart attacks are common, an

essentially random enrollment procedure yielded sufficient cases and specific advance provision could be made for at least limited control of confounding. In addition, among 25 000 admissions there were also sufficient primary diagnoses independent of aspirin use from which to select a valid control series. The “special study” confirmed the inverse association,¹⁸ the validity of which has since been established beyond all reasonable doubt.²⁰ Second, the “special study” demonstrated that it was feasible to use the case-control approach as a surveillance methodology (see below). Third, the focus had now shifted in the direction of monitoring drug use in the population at large in relation to the occurrence of serious diseases—which is the principal concern with regard to issues of drug safety.

Experience also revealed that there were drawbacks to the “special study,” and a critical review of that experience was the final and most important factor to influence the development of CCS. It is convenient to commence that review with the testing of hypotheses proposed by others.

- (i) *Hypothesis testing.* Apart from confirming the aspirin/myocardial infarction association, the “special study” confirmed other hypotheses, such as an increased risk of major upper gastrointestinal bleeding among aspirin users,²¹ an increased risk of venous thromboembolism among oral contraceptive users,²² and a decreased risk of benign breast disease among oral contraceptive users.²² However, for the purpose of rigorous evaluation, the diagnostic data were insufficiently precise (e.g., histology of benign breast disease). Further, information on important confounders either was not recorded at all (e.g., prior history of gastrointestinal bleeding, or of thrombosis, or of breast biopsy), or it was recorded in insufficient detail. The lessons: more detailed and precise diagnostic data are needed; detailed medical history is essential; and cast the widest possible net for potential confounders.
- (ii) *Hypothesis generation and in-depth evaluation.* Previously unsuspected associations were identified, and then documented in depth:

for example, a reduced risk of functional ovarian cysts among oral contraceptive users.²³ However, a related effect, that oral contraceptive use reduces the risk of ovarian cancer,²⁴ went undetected because too few cases were enrolled, and because histories of long discontinued use were not recorded. The lessons: selectively enroll diseases of concern, and record lifetime histories of drug use. Later, when CCS carried out these measures, the negative association between oral contraceptives and ovarian cancer was confirmed.²⁵

Another association missed in the “special study” was the increased risk of endometrial cancer among estrogen users.¹¹ Again, that association was subsequently evident in the CCS data,^{26,27} within which it was possible not only to confirm what others had reported, but also, on in-depth analysis, to strengthen the evidence to suggest causality (see “Strengths”). Had the association not been known, CCS would have discovered it, and validated it. The lessons: to accrue all but the most common diagnoses, selective procedures must be used; lifetime drug histories are needed; and to validate and extend observations, large numbers are needed.

One finding, the association between the use of rauwolfia alkaloids and breast cancer,²⁸ proved to be so disastrously wrong as to cast doubt in the minds of some on the concept of hypothesis generation itself and, indeed, even on the validity of case-control methods.²⁹ In essence, the hypothesis-generating data were as follows: among 150 cases of breast cancer, 11 (7.3%) had used rauwolfia in the three months before admission, whereas the exposure rate among controls was 2.2%, for a relative risk estimate of 3.5 ($p = 0.0007$). Eventually, after a great many studies, the weight of the evidence suggested that the rauwolfia alkaloids do not increase the risk of breast cancer.³⁰ In retrospect, the original data were not valid for the following reasons.

- (a) In the course of multiple comparisons, “significant” associations are sometimes bound to arise by chance. The question of how to distinguish between chance and

causation has been much debated, and the rauwolfia/breast cancer findings stimulated the debate afresh.^{29–32} As a practical matter, however, if we assume the absence of bias or confounding, chance associations should not be replicated if the same methods are used to collect further data. If, however, there are repeated replications, a causal inference becomes more tenable.

There is also the related issue of biological plausibility. When a relatively invariate association is observed in many studies, preferably based on different methods, such plausibility can play an important role, some would argue an essential one, in the final step of reaching a causal inference. However, when an association is first identified in the course of multiple comparisons, that criterion may be of little or no help in distinguishing between chance and causation, because it is almost always possible to invoke selective biological evidence either to support an association, or not to support it. Again, regardless of biological considerations, the first requirement after identifying an unexpected association must be to re-identify it.

The lesson: try to replicate unexpected findings by collecting more data. Even then, treat your own findings with suspicion until they have been independently confirmed in other studies.

- (b) Since it eventually transpired that there is no increase in risk, the implications for a surveillance system of null findings (i.e., drug safety) must be considered. A drug surveillance system should have the capacity not only to detect associations, but also to rule them out. It is an unfortunate but inescapable statistical fact that documentation of the null within narrow confidence limits requires much larger numbers than the documentation of strongly positive associations. The lesson: enroll sufficient cases to document drug safety, on the assumption that the null

hypothesis may be the “correct” one. Subsequently, after CCS was established and more than 1000 cases of breast cancer had been studied, even small associations with rauwolfia use were ruled out.³³ The original association was not replicated, and the strong likelihood is that it was indeed due to chance.

- (c) It is impossible to stipulate the exact time of onset of breast cancer. For those currently exposed, one can only be confident that the exposure came first if it lasted many years. For that purpose there were insufficient data. For the same reason, and in order to explore latent interval or duration effects, information on discontinued rauwolfia use was needed. Such information was also needed in order to distinguish between possible cancer initiation and promotion. The lessons: collect lifetime histories of drug exposure; define the exposure so that it unequivocally antedates the disease; study sufficient cases not only to test overall hypotheses, but also to test additional ones that might hold under causal assumptions (e.g., duration, dose–response, latent interval, or promotional effects).
- (d) The study recorded data on confounders relevant mainly to the risk of myocardial infarction in relation to aspirin exposure. Virtually no information on breast cancer risk factors (e.g., parity) was collected, and confounding could not be controlled. For the same reason, it was not possible to evaluate consistency within relevant strata. In any event, with only 11 exposed cases such an evaluation would not have been feasible; still less would it have been feasible to attain statistical stability within strata. The lessons: cast a wide a net for confounders likely to be relevant to a wide range of hypotheses; for validation purposes, enroll larger numbers of cases than are needed simply to attain overall statistical significance.

In summary, the development of CCS was influenced by conceptual formulations, by early experience with drug surveillance, and by the contemporary development and refinement of case-control methods.

DESCRIPTION

CCS is conducted in full compliance with the Helsinki principles, and with NIH guidelines for human research. In addition, in all grant and contract arrangements, whether with government, industry, or others, strict provision is made for academic independence. The data collected in CCS are the exclusive intellectual property of the investigators, who are solely responsible for decisions concerning publication, or dissemination of information in other ways (e.g., reports to regulatory agencies or manufacturers).

Case-control studies can vary in their specificity (Figure 13.1). In an entirely specific study, a single exposure in relation to a single outcome is at issue; such a study is largely a conceptual abstraction, with little applicability in real life. In a semi-specific study, multiple exposures in relation to a

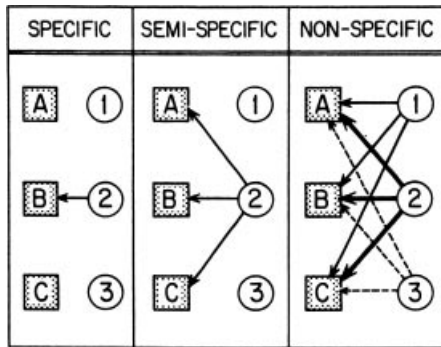


Figure 13.1. Types of case-control study (see text). In a specific study a single exposure (drug) is evaluated in relation to a single outcome (disease). In a semi-specific study multiple exposures are evaluated in relation to a single outcome. In a nonspecific study multiple exposures are evaluated in relation to multiple outcomes. The arrows point from outcome to exposure because the cases are identified first, and the exposures are then determined. (For simplicity, controls are omitted from the figure.)

single outcome are at issue. *Ad hoc* case-control studies conform to the semi-specific model. The unique contribution of CCS is that it is an extension to a nonspecific model, in which a wide range of exposures are studied in relation to a wide range of outcomes, within a single database.

METHODS

Nurse interviewers are stationed in selected hospitals, where they identify cases according to a priority list of diseases of special interest (Figure 13.2). Only incident cases diagnosed no more than a year previously are enrolled. Patients with a wide range of conditions other than those on the priority lists are also interviewed. The priority list is changed from time to time, partly in response to what is currently of concern, and partly based on review of the accumulated data. If sufficient cases of a given disease have been enrolled, that disease is removed from the list. Alternatively, from time to time new diseases are added. The nurses use sources such as admission lists, ward files, operation logs, and pathology logs to identify cases.

For some diseases, even the use of priority lists may yield insufficient cases, and it may be

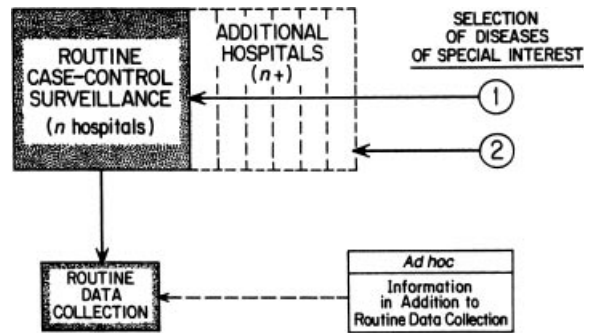


Figure 13.2. Case-control surveillance (see text). Surveillance is routinely conducted in hospitals (n) in which priority lists guide the selection of diseases of special interest; other diseases not on the priority lists are also selected. If sufficient priority list cases cannot be accrued, additional hospitals ($n+$) are added. Trained nurse interviewers administer standard questionnaires. If a particular research question necessitates the collection of additional information, the routine data collection is augmented on an *ad hoc* basis.

necessary to augment the routine data collection by special means (Figure 13.2). In recent years, CCS has focused most particularly on cancer risk, and has enrolled large numbers (sometimes thousands) of cases (see Table 13.1): to do so, appropriate hospitals, and within them, when there have been concerns about specific cancers, appropriate departments, have been added to the system (e.g., cases of malignant melanoma were enrolled from a melanoma clinic.^{34,35}

A standard data collection form is used for routine surveillance. It has five parts:

1. *Demographic data.* The usual information is recorded. Examples include age, sex, race, weight, height, marital status, and religion at birth (these data, of course, are also potentially confounding variables).
2. *Lifetime histories of drug use.* The patients are questioned about the use of drugs in the year before admission, or at any earlier time, according to over 40 indication categories. The categories are not mutually exclusive or hierarchical. The wording is colloquial, and intended to prompt memory. Examples are: pain/headache/backache; blood pressure; change of life/menopause/hot flushes; sleeping medications; tremors/seizures/anticonvulsants. The emphasis is on the regular use of drugs, but all episodes of reported use are recorded. With the exception of oral contraceptives and conjugated estrogens, doses are not recorded. For analytical purposes, short episodes of use that took place within days or weeks of admission are considered to be sufficiently precise for the evaluation of acute drug effects (e.g., aspirin and gastrointestinal bleeding³⁶). However, it is unlikely that brief or occasional drug use in the distant past can be reliably remembered (see Chapter 39). Thus, although it is operationally convenient to record all reported use, as a general rule drug use that occurred more than a year previously is only analyzed if it was regular, and if it lasted for months or years. In addition, if there is any question as to the time sequence between exposure and outcome, only regular and long term use that clearly antedated the onset of the disease is analyzed
3. *Confounding factors.* The possibility of confounding cannot ever be entirely eliminated in nonexperimental research, and sources of confounding can be multiple, subtle, and complex.³⁸ Thus, relative to an *ad hoc* study designed to test a specific hypothesis, a multi-purpose system can seldom make the equivalent advance provision for the recording of all confounding factors relevant to hypotheses still to be generated and tested at a future date. Despite these intrinsic limitations, however, it is nevertheless feasible to record a sufficiently wide range of potential confounders to meet most of the requirements of routine surveillance (Figure 13.2). The routinely recorded potentially confounding factors can be broadly divided into three categories: (i) the standard demographic variables, mentioned above, (ii) comprehensive medical histories: all major illnesses and episodes of major surgery—and, for female patients, full gynecological and obstetrical histories, and (iii) environmental and lifestyle factors that may be correlates both of drug use and disease risk: smoking habits, alcohol, coffee and tea consumption; years of education; physical activity; and frequency of health care utilization. In addition, on those occasions when they are not met, *ad hoc* information on relevant risk factors can be added to the routine data collection (Figure 13.2). For example, in order to evaluate oral contraceptives and malignant melanoma risk,³⁵ information on sun exposure was added as a potential confounder.
4. *Genetic data.* It has long been known that genetic factors play a role in determining the occurrence of certain drug effects.³⁹ Case-control surveillance has focused on cancer risk, and for that purpose the collection of standard data on biomarkers of genetic susceptibility is desirable. Recently it has been shown that DNA can reliably be extracted from buccal swabs, and stored.⁴⁰ The method is inexpensive and feasible on a large scale, and it has been added to CCS. Stored DNA will thus be available for testing future hypotheses.

(e.g., a reduced risk of cancer of the colon and rectum among aspirin users³⁷).

5. *Diagnostic data.* The discharge summary, and, when appropriate, the pathology report (e.g., cancer), is obtained. The primary diagnosis, and up to five secondary diagnoses, are recorded.

HYPOTHESIS GENERATION

The completed data are stored in master computer files. At regular intervals, multiple comparisons are carried out. Drug exposure frequencies for each disease are compared with the corresponding frequencies among the remaining patients, after adjustment for age, sex, and geographic region. The associations that emerge may be obviously noncausal (e.g., trinitroglycerin and myocardial infarction), tentative and requiring further replication before they are considered seriously, or eligible for in-depth analysis. Unexpected associations observed for the first time have to be replicated before they are considered further. Three examples are useful in illustrating how the screens are used:

1. *Alcohol and breast cancer.*⁴¹ When only a few hundred cases of breast cancer had been studied, an unexpected association with alcohol intake was identified; it was judged to be too far-fetched to deserve more detailed scrutiny. The association was then repeatedly replicated as more data were collected. It was eventually analyzed in depth, and published as a hypothesis. Even then, because the association was of relatively low magnitude, confounding could not be ruled out, and the hypothesis was stated tentatively. The association has since been confirmed in some studies, but not in others,⁴² including an *ad hoc* study which we undertook ourselves.⁴³ At present it is by no means certain that alcohol increases the risk of breast cancer.⁴²
2. *Vasectomy and prostatic cancer.*⁴⁴ One association identified in the multiple comparisons was an apparent increase in the risk of prostatic cancer among men who had undergone vasectomy. In-depth analysis was deferred pending replication. The association was then reported independently by another group.⁴⁵ At that

point, the data were analyzed in depth, and published as a hypothesis.⁴⁴ In an editorial comment that accompanied publication of the hypothesis, Guess⁴⁶ reasoned that it was by no means certain that the association is causal.

3. *Estrogens and ovarian cancer.* During the first two years of CCS, the multiple comparisons resulted in a statistically significant positive association between estrogen use and ovarian cancer. However, as more data were collected, the association was not replicated. Eventually, after much more data had been collected, the relationship was analyzed in depth, mainly because a positive association with endometrioid ovarian carcinoma had been published.⁴⁷ In the CCS data the risk of ovarian cancer, and of endometrioid carcinoma in particular, did not appear to be increased by estrogen use.⁴⁸ In the absence of an ongoing data collection system, the original association might well have been published, and subsequently have been shown to be disastrously wrong.

IN-DEPTH ANALYSIS

Whether a hypothesis is derived from the multiple comparisons, or proposed by others, the procedures for in-depth analysis are the same. In all instances a spin-off tape is abstracted from the master files. To do so, the first step is to specify the hypothesis precisely; and then to specify the cases, controls (including the number, the acceptable diagnoses, and if relevant, the matching criteria), and the potentially confounding variables. The data are then ready for analysis. When there appear to be relatively few confounders (which is rare), the analysis may require little more than minimal adjustment for confounding by factors such as age. For more complex situations with multiple confounders (which is common), multivariate methods are used.

STRENGTHS

Since its foundation in 1975, CCS has enrolled some 82 000 patients, mainly in Massachusetts, New York, Pennsylvania, and Maryland. Over 70

publications have appeared in medical journals, book chapters, symposia, and the like. A full publication list is provided in the appendix.

STATISTICAL POWER

Table 13.1 lists the primary diagnoses for which the accrual has exceeded 100 cases. Table 13.2 lists the drugs to which more than 1% of the population have been exposed at some time in their lives; in the range of 0.5–2%, several hundred drugs have been used (data not given). Table 13.3 classifies ever-use of drugs according to 50 categories, with exposure rates ranging from 1 to 60%. Table 13.4 gives estimates of numbers of cases required to document statistically significant relative risks of 0.5, 2.0, 3.0, 5.0, and 10, according to drug exposure rates of 0.5, 5, 10, and 25% (see also Appendix A of this book). Table 13.5 gives, for the same rates of drug exposure, the numbers of cases required to document an upper two-sided 95% confidence limit of <2.0 , assuming the relative risk point estimate is 1.0. Cross-referral among the five tables indicates in broad quantitative terms the statistical power of CCS. It is not possible to consider the tables exhaustively, but a few salient points will be made.

Case-control methods are most efficient in the evaluation of relatively common exposures. When, in addition, diseases can be enrolled in very large numbers (thousands), it is possible to document modestly elevated risks (≥ 2.0) based on control exposure rates as low as 0.5%. For higher exposure rates power is considerably greater. Alternatively, if there is no apparent increase in risk (relative risk ≈ 1.0) exposure rates of 1–2% make it possible to “rule out” material increases in risk (upper confidence interval, <2.0). With large numbers, it is also possible to identify protective effects for commonly used drugs (exposure rate, $>2\%$). For exposure rates much below 0.5%, it is beyond the capacity of the system to document safety; positive associations, if strong, can still be documented, but power declines.

For diseases that can be enrolled in lesser but still adequate numbers (hundreds), for drug exposure rates in excess of 1–5% it is still possible

Table 13.1. Primary diagnoses in case-control surveillance, 1975–1998 (numbers rounded off)

Disease	No.
Breast cancer	8076
Fractures/traumatic injury	3500
Lung cancer	1900
Malignant melanoma	1830
Cholecystitis	1800
Colon cancer	1800
Ovarian cyst	1500
Myocardial infarction	1400
Ovarian cancer	1200
Endometrial cancer	1200
Prostate cancer	1185
Appendicitis	1100
Rectal cancer	1100
Herniated lumbar disc	1000
Cholelithiasis	1000
Pelvic inflammatory disease	850
Inguinal hernia	790
Venous thromboembolism	600
Fibrocystic breast disease	600
Leukemia	600
Bladder cancer	590
Pancreatic cancer	549
Non-Hodgkin's lymphoma	530
Ectopic pregnancy	490
Renal calculus	460
Renal cancer	430
Osteoporotic fractures	430
Testicular cancer	410
Diverticulitis	400
Stomach cancer	370
Spontaneous abortion	360
Hodgkin's disease	360
Cataract	330
Pancreatitis	310
Detached retina	290
Duodenal ulcer	280
Major upper gastrointestinal bleeding	240
Esophageal cancer	230
Gastric ulcer	210
Bone cancer	180
Hemorrhoids	170
Pulmonary embolism	140
Salpingitis	140
Crohn's disease	140
Oral cancer	130

to document increased risks of some twofold or greater. For null estimates, however, it is not possible to exclude modest increases in risk, unless the exposure rates are reasonably high ($>5\%$).

Table 13.2. Ever-use of drugs in case-control surveillance, 1975-1998

Drug name	%
Aspirin	34
Acetaminophen	31
Benzodiazepines	22
Ascorbic acid	17
Tetracycline	10
Ibuprofen	10
Ampicillin	9
Hydrochlorothiazide	7
Erythromycin	7
Furosemide	6
Prednisone	6
Bufferin®	5
Propoxyphene HCl	5
Chlordiazepoxide	5
Propranolol	5
Percodan®	5
Acetaminophen with codeine	5
Methyl dopa	4
Codeine	4
Sulfisoxazole	4
Contact®	4
Ovral®	4
Norinyl®	4
Conjugated estrogens, 0.625 mg	4
Diphenoxylate HCl/atropine	3
Warfarin sodium	3
Nitroglycerin	3
Milk of magnesia	3
Indomethacin	3
Chlorpheniramine maleate	3
Flurazepam HCl	3
Cephalexin monohydrate	3
Vitamin A	3
Cimetidine	3
Levothyroxine sodium	3
Heparin sodium	3
Diphenhydramine HCl	3
Fiorinal®	2
Digoxin	2
Meperidine HCl	2
Chlorothiazide	2
IUD, Lippes loop	2
Amitriptyline HCl	2
Methocarbamol	2
Nystatin	2
Metronidazole	2
Prochlorperazine	2
Tetrahydrozoline HCl	2
Phenylbutazone	2
Phenytoin	2
Alka Seltzer®	2
Medroxyprogesterone acetate	2
Pseudoephedrine HCl	2
Zinc	2
Isosorbide dinitrate	2

Table 13.3. Ever-use of drugs in case-control surveillance, classified according to categories. 1975-1998

Category	%
Aspirin containing drugs	60
Vitamins/minerals	55
Iron salts	40
Acetaminophen containing drugs	39
Benzodiazepines	30
Oral contraceptives	30
Antihistamines	28
Caffeine containing drugs	28
Natural penicillins	27
Estrogens	26
Narcotic pain formulas	24
Corticosteroids	20
Sulfonamides	17
Thiazides	17
Vitamin A	16
Codeine containing drugs	13
Tetracyclines	13
Ampicillin/amoxicillin	12
Nonsteroidal anti-inflammatory drugs (excludes aspirin)	12
Calcium salts	11
Phenylpropanolamine	11
Folic acid containing vitamins	10
Spermicides	10
Barbiturates	8
Conjugated estrogens	8
β -blockers	8
Laxatives—stimulant type	8
Thyroid supplements	7
Macrolide antibiotics	7
Antifungals	7
Intrauterine devices	6
Phenothiazines	6
Phenobarbital	6
Guaifenesin	6
Nitrates	5
Potassium salts	5
H ₂ blockers	4
Docusate salts	4
Vaginal douches	4
Cephalosporins	4
Antidepressants	4
Laxatives—emollient type	3
Progestogens	3
Sulfonylureas	3
Laxatives—bulk type	3
Xanthines (excludes caffeine)	2
Spirolactone	1
Antiarrhythmics	1
Aminoglycosides	1

Table 13.4. Estimated numbers of cases required to document various relative risks (assumptions: power = 80%, $\alpha = 0.05$; control-to-case ratio = 4 : 1)

Exposure rate in controls (%)	Relative risk				
	0.5	2.0	3.0	5.0	10.0
0.5	6241	2684	838	287	83
1	3118	1350	418	148	44
2	1572	690	218	76	25
5	632	296	94	36	13
10	323	165	56	22	9
25	139	92	35	16	8

Table 13.5. Estimated numbers of cases required to document an upper 95% confidence limit (two-sided) of <2.0, assuming a "true" relative risk of 1.0 (assumptions: power = 80%, $\alpha = 0.05$; control-to-case ratio = 4 : 1)

Exposure rate in controls (%)	Number of cases
0.5	4103
1	2062
2	1041
5	430
10	227
25	109

That is, power to reject the null is still adequate, but power to document the null is more limited.

For diseases that can only be enrolled in more modest numbers (about 50–100 cases) increases in risk of the order of threefold or greater can be documented, but, assuming the true relative risk is unity, it would not be possible to rule out appreciable elevations in risk. As numbers decline further, only powerful associations with common drug exposures can be documented (for example, an increased risk of liver cancer among users of oral contraceptives, based on only 12 cases⁴⁹); and it is beyond the capacity of CCS to demonstrate safety.

Based on these projections, and on the disease frequencies in Table 13.1, CCS has the capacity to monitor a wide range of drugs commonly used in the population at large, in relation to a wide range of diseases, and to document increased risks. With large numbers CCS can document decreased risks, or it can document safety.

However, many drugs are used by appreciably less than 0.5% of the population. For the most part such drugs are not evaluable by CCS (see "Weaknesses").

To sum up, for the evaluation of commonly used drugs in relation to common diseases, which is the main public health concern, CCS has a capacity for the documentation of adversity, protection, and safety. As the diseases become less common and more difficult to accrue, or as the exposure rates decline, that capacity is correspondingly reduced. Of course, by the same token, so are the public health concerns. Even for quite rare diseases, increases in risk due to commonly used drugs can be documented.

STRENGTH OF ASSOCIATIONS

Given adequate study design, and statistical stability, when a relative risk estimate exceeds 2.0 (or is less than the inverse estimate of 0.5), as a rule it is unlikely that residual sources of bias or confounding, if identified and allowed for, would entirely eliminate it. For more modestly elevated relative risks (<2.0), or modestly reduced ones (>0.5), however, it is seldom within the resolving power of nonexperimental research to show that an apparent association is "real." This is not an issue of statistical significance: minor sources of bias or confounding can produce spurious associations of low magnitude; if the biases remain invariate the attainment of "significance" then becomes simply a matter of augmenting the sample size.³⁸ Thus, Tables 13.1–5 show that CCS can document a wide range of increases in risk of the order of twofold or greater. It is also possible to document strong protective effects (relative risk, <0.5). When diseases can be accrued in large numbers (hundreds or thousands), lesser associations can also be documented, but their interpretability becomes more difficult (e.g., alcohol and breast cancer⁴¹).

MISCLASSIFICATION OF EXPOSURE AND OUTCOME

Misclassification of exposure is present in CCS inasmuch as memory loss is unavoidable in the

reporting of drug use (see “Weaknesses”). However, in terms of specifying the time relationships between exposure and outcome, CCS has greater precision than any other surveillance methodology. Prompt access to the patients ensures that for acute effects (e.g., bleeding) the timing of drug use and of disease onset can be determined to the nearest day; for long term effects (e.g., carcinogenesis), CCS is the only surveillance system (as distinct from *ad hoc* studies) that routinely and systematically records lifetime drug exposures; and it is the only ongoing system that routinely records the use of nonprescribed drugs.

With regard to defining the outcomes, CCS is precise because of access to discharge summaries and pathology reports. The recording of medical histories also ensures that only incident cases are studied (e.g., venous thromboembolism).⁵⁰

SELECTION BIAS

The refusal rate in CCS is 4.5%, and bias due to losses can virtually be ruled out. In theory, the biased selection of cases is possible if a drug increases (or decreases) the likelihood that the disease will be fatal before hospitalization, or before interview. In practice that bias has yet to be demonstrated. Bias may also arise, however, if the likelihood of hospitalization is increased by medical surveillance (e.g., selective hospitalization for breast cancer of oral contraceptive users who undergo regular breast examinations⁵¹), and special measures may be needed to minimize it.

With regard to the selection of controls for in-depth analyses, the use of hospital controls raises the theoretical possibility of a selection bias. However, the pool of patients from which to make that selection is so large that a valid control series can almost always be defined. Occasionally, however, if the overall probability of hospitalization across a wide range of diagnoses is influenced by the exposure (e.g., alcohol, cigarettes), it may not be possible to confidently eliminate selection bias.

INFORMATION BIAS

A limitation to the validity of interview based case-control studies is their susceptibility to bias if

the cases, prompted by knowledge of the hypothesis, recall their drug exposure histories more fully than the controls. In CCS, information bias does not exist for hypothesis generation since there is no hypothesis to be aware of. When it comes to hypothesis testing, however, bias cannot be entirely ruled out (see “Weaknesses”).

CONFOUNDING

In nonexperimental research there is no validity issue more important than confounding. For drug surveillance, potential confounding is, if anything, of even greater importance because drug use is itself a health-related activity. Factors such as previous medical history and lifestyle factors profoundly influence drug use, and they profoundly influence disease risk. Most of the information required to control confounding can be obtained only by direct and prompt access to the patients. (An incomplete list of some of the more important confounders would include the following: years of education; family history of cancer; history of major illnesses; surgical history; obstetrical and gynecological history; lifetime history of nonprescription drug use; lifetime smoking history; lifetime history of alcohol and beverage intake.) A capacity to adequately control confounding is essential to drug surveillance, and we feel that CCS fulfills that requirement more fully than any other system.

CONSISTENCY

A causal inference is strengthened if the data are consistent, and statistically stable, across strata that make biological sense (e.g., age), or across technical strata relevant to validity (e.g., region). This requirement necessitates larger numbers than those required simply for overall statistical significance. Consistency also applies to dosage or duration effects, when these can reasonably be anticipated. Again, larger numbers are needed. Moreover, the analysis of effects according to duration or dose must be confined to the domain of the exposed, with the nonexposed excluded,⁵² necessitating still larger numbers: the error of including the nonexposed in testing for dose-response still occurs far

too commonly.⁵² CCS has only limited data on dosage (see “Weaknesses”); however, for the analysis of duration effects, the recording of lifetime histories of drug use is a strong advantage.

WEAKNESSES

STATISTICAL POWER

For practical purposes, it is beyond the capacity of CCS to evaluate exposure rates below 0.5%. Many drugs are used less commonly, and as a rule they can be monitored only by means of followup methods (or by adaptations, such as nested case-control studies), and only to a limited extent.

The rarity of some diseases is also a limiting factor. CCS has shown that selective enrollment is effective in accruing even quite uncommon conditions such as malignant melanoma^{34,35} or ovarian cancer.^{25,48} However, if the outcome is exceedingly rare, special ascertainment procedures covering a massive population base may be required in order to attain adequate numbers (e.g., agranulocytosis⁵³). In certain circumstances, if both the exposure and outcome are rare, but a large proportion of the outcome is due to the exposure, a causal effect may only be demonstrable by means of an *ad hoc* study (e.g., *in utero* exposure to diethylstilbestrol¹⁰). However, if only a small proportion of the outcome is due to the exposure, a causal effect may be detectable only by means of case reports (e.g., oxymetholone and peliosis hepatis/hepatic cancer⁵⁴).

STRENGTH OF ASSOCIATIONS

For relative risk estimates below 2.0, no surveillance methodology, CCS included, is capable of distinguishing between bias, confounding, and causation. Sometimes the likelihood of bias may be remote. However, it is never possible to be confident that confounding has been fully controlled, and this is more particularly a problem in the setting of surveillance systems. Even in *ad hoc* studies, the distinction between causal and non-causal associations can only occasionally be made in the context of a large amount of collateral

evidence (e.g., side-stream cigarette smoke and lung cancer⁵⁵).

MISCLASSIFICATION OF EXPOSURE AND OUTCOME

In the absence of systematic bias, the effect of misclassification is almost always to dilute or obscure associations. In interview-based case-control studies, failure to fully recall exposure, or its timing and duration, is inevitable. Memory loss can be reduced by good study design, but it cannot be eliminated. However, given adequate study design, and in particular adequate questionnaire design, validation studies have shown that there is little memory loss for recent exposures, or for long duration ones⁵⁶ (see also Chapter 39). The main problem is with the accuracy of recall of short duration exposures in the distant past. It is known, for example, that metronidazole damages DNA;⁵⁷ the drug may therefore be carcinogenic to humans, but the recall of no more than a week or two of exposure that took place many years previously is beyond the scope of CCS. However, the valid evaluation of such exposures probably is beyond the scope of any other surveillance methodology as well.

SELECTION BIAS

It is sometimes claimed that hospital-based case-control studies are limited because of potential biases in the selection of controls.⁵⁸ Sometimes that claim may be justified when insufficient attention is given to exactly which hospital controls to select (i.e., diagnoses independent of the exposure). It is not justified, however, if the control selection is both independent of the exposure, and representative of the proper study base.⁵⁹⁻⁶² Indeed, there are circumstances (e.g., when the cases represent a secondary study base)⁵⁹⁻⁶² when properly selected hospital controls may be the only valid reference group, as is generally applicable to CCS: the controls should be persons who would be admitted to the same hospitals as the cases, were they to develop the disease under study. As already mentioned, however, it may not be possible to meet the criterion of

independence when an exposure increases (or decreases) the overall probability of hospitalization across many diagnoses.

INFORMATION BIAS

When the hypothesis is known, interview-based case-control studies can seldom, if ever, demonstrate the total absence of information bias. Such bias can, however, be kept to an absolute minimum. For recent exposures, prompt access to the patients minimizes the likelihood of material memory loss; and memory loss can be further reduced by detailed attention to questionnaire design, and the use of prompts. The risk of differential recall between cases and controls may be greater for exposures in the distant past, but the likelihood of differential recall is reduced if the exposure has lasted for many years.

Validation studies have suggested that information bias, with the qualifications mentioned above, is not an intractable problem in well designed case-control studies⁵⁶ (see also Chapter 39). Nevertheless, it would considerably strengthen the validity of CCS if reported exposures could be confirmed in an independent source, such as automated prescription records. Such records, however, also have problems, such as scanty or absent information on the following: adherence to therapy; precise timing of drug use, particularly in relation to acute events; lifetime history of drug exposure; and the use of nonprescribed drugs.

CONFOUNDING

Despite elaborate provision for control of confounding, circumstances arise in CCS when that provision is incomplete. Sometimes, it may be possible to report a finding in tentative terms, with the proviso that more detailed *ad hoc* research should be undertaken. For example, an in-depth analysis of the CCS data suggested that the use of aspirin and other nonsteroidal anti-inflammatory drugs may confer protection against colonic and rectal cancer,³⁷ but confounding from certain sources (e.g., diet) was not controlled. A more detailed *ad hoc* study is now in progress. One source of confounding, and a common one in drug

epidemiology, arises when the indication for the drug is confounded with the outcome⁶³ (see also Chapter 34). In that circumstance, the confounding can seldom be measured with adequate precision, and, as amply demonstrated in the recent fenoterol/asthmatic death controversy,⁶⁴ there is no surveillance methodology that can cope with it. In that situation, a valid answer to the research question can be obtained only by means of a randomized controlled trial.⁶³

CONSISTENCY

The demonstration of a dose-response effect can be important in strengthening a causal inference. In CCS it has not proven feasible to collect dosage information: apart from problems such as variations between drugs in terms of dose formulations, schedules, and routes of administration, patients can commonly recall drug names, but not the doses. With a few exceptions, CCS cannot evaluate possible dose-response effects. On the other hand, however, CCS can evaluate duration effects (see "Strengths").

PARTICULAR APPLICATIONS

CCS is applicable to the assessment of the risk of serious diseases in relation to drug use in the population at large. CCS can generate hypotheses, and document increased risk, decreased risk, and absence of risk. When particular issues arise, CCS has the flexibility to rapidly accrue diseases of interest. Diseases so rare that they cannot be accrued in adequate numbers are beyond the scope of CCS. CCS has only limited capacity to assess uncommonly used drugs.

From the public health perspective, CCS has made major contributions to the evaluation of the health effects of supplemental female steroids, and of contraceptive methods among males and females. The diseases studied have included female cancers (breast; endometrium; ovary; hydatidiform mole; and choriocarcinoma), male cancers (prostate; testis), other cancers (large bowel; rectum; melanoma; liver); and nonmalignant conditions such as cardiovascular disease,

venous thromboembolism, pelvic inflammatory disease, gynecologic disorders, and gastrointestinal disorders. Common drugs such as analgesics, female steroid hormones, antihypertensives, diuretics, hypnotics, and tranquilizers, have been evaluated in relation to a wide range of outcomes. CCS has also made contributions to epidemiological methods, and to the evaluation of risk factors other than drugs (e.g., cigarettes; coffee; alcohol; vasectomy). Increased risks, decreased risks, and safety have all been documented, and some of the findings have been new discoveries.

HYPOTHESIS GENERATION AND TESTING

It is encouraging (but not surprising) to note that most of the multiple comparisons undertaken in CCS have generally favored drug safety. By definition, null effects do not result in hypothesis generation. Yet, although in number and importance they far exceed deviations from the null, as a rule evidence of safety is only published if a suspicion to the contrary has been raised (see below: "Testing of hypotheses proposed by others"). We consider the capacity to document safety to be one aspect of conducting multiple comparisons, and an important application of CCS.

The capacity of CCS to document drug safety is especially important. Recently, for example, based on data from a followup study^{65,66} it was claimed that calcium channel blockers increase the overall risk of cancer, and inhibition of apoptosis was proposed as the biological basis for the association. In the CCS data⁶⁷ an increased risk could be ruled out with close statistical confidence. There was also sufficient power to rule out increased risks for each of the specific neoplasms included in the original report.^{65,66} In that report the data for the individual neoplasms were sparse, and statistically unstable. To give another example, a purported association between acetaminophen use and cancer of the kidney that was based on inadequate data⁶⁸ could be rebutted.⁶⁹

Some hypotheses have been generated. One of them is an apparent increased risk of cholecystitis among users of thiazide diuretics.⁷⁰ Another association is an apparent increased risk of

choriocarcinoma among women previously exposed to oral contraceptives.⁷¹ A more controversial relationship is the association between alcohol use and breast cancer.^{42,43} A finding of interest, but not a drug-related one, is an association between vasectomy and prostatic cancer.⁴⁴ A reduced risk uncovered by CCS is the inverse association between the use of nonsteroidal anti-inflammatory drugs and large bowel cancer.³⁹ These associations have received varying degrees of confirmation in other studies,^{42,72,73} and our own *ad hoc* research has recently provided confirmatory evidence to support the apparent reduction of the risk of cancer of the large bowel among nonsteroidal anti-inflammatory drug users.⁷⁴

TESTING OF HYPOTHESES PROPOSED BY OTHERS

CCS has sometimes confirmed and sometimes rebutted hypotheses proposed by others. Rebutted hypotheses have included rauwolfia and breast cancer,³³ hydralazine and various cancers,⁷⁵ diazepam and breast cancer,⁷⁶ oral contraceptives and malignant melanoma,^{34,35} and most recently calcium channel blockers and all-cancer risk.⁶⁷

A large number of associations have been confirmed, and sometimes the CCS data could also be used to extend them, and enhance their validity. For example, in two sequential studies, CCS first confirmed²⁶ that estrogen use increases the risk of endometrial cancer, and in addition showed that the increased risk persisted after discontinuation of use. This observation discounted the argument that current estrogen use causes uterine bleeding, and hence the selective diagnosis of otherwise silent endometrial cancer.⁷⁷ Then, after more data had been collected, CCS showed that estrogens increase the risk of both invasive and noninvasive cancer,²⁷ that the elevated risk persists for at least 10 years following discontinuation of use, and that, even after such long discontinued use, the risk was duration-related.

Another example of confirmation and extension of a previously reported association is the increased risk of pelvic inflammatory disease among

women using intrauterine devices.^{78,79} CCS data confirmed the association, and suggested in addition that the risk was duration-related. It also suggested that the risk may have been higher for the Dalkon Shield, and lower for copper containing devices.

Two reduced risks confirmed by CCS are the associations of oral contraceptive use with endometrial cancer⁸⁰ and ovarian cancer.²⁵ CCS extended the original observations by suggesting that the protective effects are duration-related, and that they persist for many years after oral contraceptives are discontinued.

CASE-CONTROL SURVEILLANCE OF BIRTH DEFECTS

CCS was developed to assess drug-related risks in adult populations. In conceptual terms, however, the methodology is also applicable to other populations, and it has been applied to the monitoring of *in utero* drug exposure in relation to the risk of birth defects. As with CCS, the system has shown the capacity to generate and test hypotheses, to document positive and inverse associations, to document safety, and to be flexible in focusing on outcomes of concern, as needed. Methodological issues pertinent to the surveillance of birth defects are considered in Chapter 42.

THE FUTURE

New drugs cannot be said to have been adequately monitored until large populations have been exposed, and until those drugs have been monitored in relation to the risk of commonly occurring diseases. When applicable, the evaluations must include long term drug use—and, for cancer in particular, long latent intervals. Thus, we believe that CCS should continue indefinitely for as long as new drugs continue to be introduced. Even for well established drugs, CCS still needs to accrue diseases not yet accrued in sufficient numbers. Some of the oldest drugs continue to produce surprises. Aspirin, of course, is the prime example: increased risks (e.g., Reye syndrome⁸¹) continue to be identified, as do reduced risks (e.g.,

heart attack,¹⁸⁻²⁰ large bowel cancer³⁷) In short, all marketed drugs must be subjected to continuing surveillance for as long as they exist.

CCS cannot monitor exceedingly rare diseases, but the possibility that the methods can be suitably modified in order to do so is being explored. Wiholm⁸² has reviewed the reasons, worldwide, why governments have banned marketed drugs: 70% of the reasons can be accounted for by a relatively short list of rare diseases known to be caused by many drugs: for example, blood dyscrasias (e.g., agranulocytosis); hepatic disorders (e.g., hepatic necrosis); renal disorders (e.g., acute renal failure); and dermatological disorders (e.g., toxic epidermal necrolysis). For CCS to be applicable to the evaluation of the drug-related risks associated with these diseases, ascertainment mechanisms capable of identifying cases from base populations of many millions would be needed. That such an approach can be feasible and informative has already been shown in large scale *ad hoc* studies of blood dyscrasias.^{53,83,84} An international case-control study of life-threatening skin disorders also supports the feasibility of this approach (unpublished data), as does a recently commenced study of anaphylactoid reactions (unpublished data).

Our experience with two of these diseases (agranulocytosis and aplastic anemia)⁵³ indicates that they pose unique methodological problems, particularly with regard to biased misclassification of the timing of drug exposure and disease onset, and confounding. An example of misclassification is the association of antimicrobial use with agranulocytosis.^{53,83} Agranulocytosis typically presents with sore throat, chills, and fever. Without accurate information on timing of exposure to the nearest day, and of disease onset, it is impossible to determine whether the antimicrobial caused the disease, or was administered to treat the symptoms. Similarly, for the valid study of anaphylactoid reactions, the timing information needs to be accurate to the nearest hour. Control of confounding is especially important because the diseases are caused by multiple drugs, and cases have commonly been exposed to more than one drug.⁵³ Based on a large experience with agranulocytosis, aplastic anemia, thrombocytopenic

purpura,⁸⁴ toxic epidermal necrolysis, Stevens–Johnson syndrome, and anaphylactoid reactions we judge that most of the disorders known to be caused by multiple drugs cannot be adequately monitored by any means other than direct access to the patients.

Our experience after more than two decades of CCS has confirmed our initial forecast^{3,63,85} that surveillance can seldom fully substitute for *ad hoc* research, and that it must be complemented by such research. For example, the positive association between oral contraceptive use and choriocarcinoma,⁷¹ and the inverse association between nonsteroidal anti-inflammatory drugs and cancer of the large bowel,³⁷ have been investigated,⁷⁵ or are currently being investigated, in *ad hoc* studies that are more detailed than was feasible in CCS. There are also situations so unique that they can never be covered by surveillance: for example, the association between *in utero* exposure to diethylstilbestrol and vaginal cancer.¹⁰ We consider the capacity to mount *ad hoc* studies to be an essential component of drug surveillance itself.

One goal, not yet realized, must be to overcome the single major weakness of case–control methodology, its susceptibility to information bias. In conceptual terms, if CCS were to be conducted in populations whose drug prescription data are automated,^{86,87} that weakness could be overcome, at least in part. Such a system would also offer the advantage of adequate control of confounding, and greater precision in the definition of outcomes—the Achilles’ heel of automated databases. Nested case–control methodology would also offer considerable advantages in terms of efficiency. Such an approach would thus combine the best features of case–control methodology and automated databases.

Finally, greater scientific excellence must be a future goal for CCS, as it must be for all of those engaged in the epidemiological evaluation of drug safety or adversity. The ultimate, if elusive, goal should be to avoid the experience of being disastrously wrong. Issues of drug safety, or adversity, lie at the intersection of conflicting pressures from politicians, regulators, manufacturers, public interest groups, the medical profession, academic peers, and others. We reject the

argument that because of the pressures, there is sometimes a need for quick and cheap answers, if only approximate ones.^{86–92} Such answers are seldom either quickly available, or cheap—and they may be erroneous, as occurred, for example, when inadequate data, inadequately and hastily analyzed, produced spuriously low estimates of the risk of gastrointestinal bleeding due to the use of piroxicam.^{93–95} It is the responsibility of scientists to resist the pressures, and to insist on proper standards of evidence. Scientifically unsatisfactory evidence “.... is not considered acceptable for evaluating efficacy and it ought not to be considered acceptable for evaluating adversity”.⁸⁷ Good science should be at the heart of drug epidemiology

ACKNOWLEDGMENTS

Case–control surveillance was first conceived by the late Dennis Slone and myself in 1975. I thank David Kaufman, Allen Mitchell, and Lynn Rosenberg for suggestions made in the course of preparing this chapter. However, I take sole responsibility for the contents.

Since its foundation, case–control surveillance has been supported mainly by contracts and grants from the United States Food and Drug Administration, and by a grant from the National Cancer Institute (R37 CA45762. principal investigator: L. Rosenberg). Additional support has been received from Astra Pharmaceutical Products Inc.; Bayer AG; Ciba–Geigy Corporation; Hoechst AG; Hoffmann–La Roche, Inc.; Marion Merrell Dow Inc.; McNeil Consumer Products Company; McNeil Pharmaceutical; National Coffee Association of USA, Inc.; Merck and Company, Inc.; Merkle; Pfizer Pharmaceuticals; G. D. Searle and Company; The Upjohn Company; and Sterling Winthrop Inc.

APPENDIX: CASE–CONTROL SURVEILLANCE. PUBLICATIONS 1975–1998

1. Slone D, Shapiro S, Miettinen OS. Case–control surveillance of serious illnesses attributable to

- ambulatory drug use. In: Colombo F, Shapiro S, Slone D, Tognoni G, eds, *Epidemiological Evaluation of Drugs*. Amsterdam: Elsevier-North-Holland, 1977; 59-70.
2. Miettinen OS, Slone D, Shapiro S. Current problems in drug-related epidemiologic research. In: Colombo F, Shapiro S, Slone D, Tognoni G, eds, *Epidemiological Evaluation of Drugs*. Amsterdam: Elsevier-North-Holland, 1977; 295-303.
 3. Shapiro S, Slone D. Case-control surveillance. In: Gross FH, Inman HW, eds, *Drug Monitoring*. London: Academic, 1977; 33-48.
 4. Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. A policy perspective. *Ann Intern Med* 1979; **90**: 257-61.
 5. Rosenberg L, Shapiro S, Kaufman DW, Slone D, Miettinen OS, Stolley PD. Patterns and determinants of conjugated estrogen use. *Am J Epidemiol* 1979; **109**: 676-86.
 6. Kaufman DW, Shapiro S, Rosenberg L, et al. Intrauterine contraceptive device use and pelvic inflammatory disease. *Am J Obstet Gynecol* 1980; **136**: 159-62.
 7. Kaufman DW, Slone D, Rosenberg L, Miettinen OS, Shapiro S. Cigarette smoking and age at natural menopause. *Am J Public Health* 1980; **70**: 420-2.
 8. Rosenberg L, Slone D, Shapiro S, Kaufman DW, Stolley PD, Miettinen OS. Noncontraceptive estrogens and myocardial infarction in young women. *J Am Med Assoc* 1980; **244**: 339-42.
 9. Shapiro S, Kaufman DW, Slone D, et al. Recent and past use of conjugated estrogens in relation to adenocarcinoma of the endometrium. *New Engl J Med* 1980; **303**: 485-9.
 10. Rosenberg L, Shapiro S, Slone D, Kaufman DW, Miettinen OS, Stolley PD. Thiazides and acute cholecystitis. *New Engl J Med* 1980; **303**: 546-8.
 11. Shapiro S, Kaufman DW, Slone D, et al. Use of thyroid supplements in relation to breast cancer. *J Am Med Assoc* 1980; **244**: 1685-7.
 12. Hartz SC, Shapiro S, Kaufman DW, Rosenberg L, Slone D. Factors associated with oral contraceptive use. *Am J Public Health* 1980; **70**: 1105-8.
 13. Kaufman DW, Shapiro S, Slone D, et al. Decreased risk of endometrial cancer among oral contraceptive users. *New Engl J Med* 1980; **303**: 1045-7.
 14. Rosenberg L, Slone D, Shapiro S, et al. Breast cancer and alcoholic-beverage consumption. *Lancet* 1982; **1**: 267-70.
 15. Kaufman DW, Shapiro S, Slone D, et al. Diazepam and the risk of breast cancer. *Lancet* 1982; **1**: 537-9.
 16. Rosenberg L, Shapiro S, Slone D, et al. Epithelial ovarian cancer and the use of combination oral contraceptives. *J Am Med Assoc* 1982; **247**: 3210-12.
 17. Helmrich SP, Shapiro S, Rosenberg L, et al. Risk factors for breast cancer. *Am J Epidemiol* 1983; **117**: 35-45.
 18. Shapiro S, Slone D, LaPorte JR. El metodo caso-control en farmacovigilancia. In: Laporte JR, Tognoni G, eds, *Principios de Epidemiologia del Medicamento*. Barcelona: Salvat Editores, 1983; 199-216.
 19. Kaufman DW, Watson J, Rosenberg L, et al. The effect of different types of intrauterine device upon the risk of pelvic inflammatory disease. *J Am Med Assoc* 1983; **250**: 759-62.
 20. Rosenberg L, Schwingl PJ, Kaufman DW, et al. Breast cancer and cigarette smoking. *New Engl J Med* 1984; **310**: 92-4.
 21. Rosenberg L, Miller DR, Kaufman DW, et al. Breast cancer and oral-contraceptive use. *Am J Epidemiol* 1984; **119**: 167-76.
 22. Helmrich SP, Rosenberg L, Kaufman DW, et al. Lack of an elevated risk of malignant melanoma in relation to oral contraceptive use. *J Natl Cancer Inst* 1984; **72**: 617-20.
 23. Shapiro S. The epidemiological evaluation of drugs. *Acta Med Scand Suppl* 1984; **683**: 23-7.
 24. Shapiro S, Parsells JL, Rosenberg L, Kaufman DW, Stolley PD, Schottenfeld D. Risk of breast cancer in relation to the use of rauwolfia alkaloids. *Eur J Clin Pharmacol* 1984; **26**: 143-6.
 25. Kaufman DW, Miller DR, Rosenberg L, et al. Noncontraceptive estrogen use and the risk of breast cancer. *J Am Med Assoc* 1984; **252**: 63-7.
 26. Rosenberg L, Miller DR, Helmrich SP, Kaufman DW, Shapiro S. Breast cancer and coffee drinking. In: *Banbury Report 17: Coffee and Health*. New York: Cold Spring Harbor Laboratory, 1984; 189-95.
 27. Miller DR, Rosenberg L, Helmrich SP, Kaufman DW, Shapiro S. Ovarian cancer and coffee drinking. In: *Banbury Report 17: Coffee and Health*. New York: Cold Spring Harbor Laboratory, 1984; 157-65.
 28. Lesko SM, Rosenberg L, Kaufman DW, et al. Cigarette smoking and the risk of endometrial cancer. *New Engl J Med* 1985; **313**: 593-6.
 29. Rosenberg L, Miller DR, Helmrich SP, et al. Breast cancer and the consumption of coffee. *Am J Epidemiol* 1985; **122**: 391-9.
 30. Lesko SM, Kaufman DW, Rosenberg L, et al. Evidence of increased risk of Crohn's disease in oral contraceptive users. *Gastroenterology* 1985; **89**: 1046-9.
 31. Shapiro S, Parsells JL, Rosenberg L, et al. Risk of localized and widespread endometrial cancer in relation to recent and discontinued use of conjugated estrogens. *New Engl J Med* 1985; **313**: 969-72.
 32. Rosenberg L. Post-marketing surveillance. The approach of the Drug Epidemiology Unit. *Drug Inf J* 1985; **19**: 263-8.
 33. Miller DR, Rosenberg L, Kaufman DW, Schottenfeld D, Stolley PD, Shapiro S. Breast cancer risk in

- relation to early oral contraceptive use. *Obstet Gynecol* 1986; **68**: 863–8.
34. Shapiro S, Dunne J, Kaufman D, eds., *Monitoring and Assessment of Adverse Drug Effects*. Geneva: CIOMS/WHO, 1986.
 35. Helmrich SP, Rosenberg L, Kaufman DW, Strom B, Shapiro S. Venous thromboembolism in relation to oral contraceptive use. *Obstet Gynecol* 1987; **69**: 91–5.
 36. Kaufman DW, Kelly JP, Rosenberg L, Stolley PD, Schottenfeld D, Shapiro S. Hydralazine and breast cancer. *J Natl Cancer Inst* 1987; **78**: 243–6.
 37. Schatzkin A, Palmer JR, Rosenberg L, et al. Risk factors for breast cancer in black women. *J Natl Cancer Inst* 1987; **78**: 213–17.
 38. Miller DR, Rosenberg L, Kaufman DW, et al. Epithelial ovarian cancer and coffee drinking. *Int J Epidemiol* 1987; **16**: 13–17.
 39. Shapiro S. Reasons for the successes and failures of specific models in drug epidemiology. In: Kewitz H, Roots I, Voigt K, eds, *Epidemiological Concepts in Clinical Pharmacology*. New York: Springer, 1987; 11–22.
 40. Rosenberg L, Lesko SM. Cigarette smoking and endometrial cancer. In: Rosenberg MJ, ed., *Smoking and Reproductive Health*. Littleton, MA: PSG, 1987; 160–6.
 41. Shapiro S. Evolution of drug monitoring. In: Shapiro S, Dunne J, Kaufman D, eds, *Monitoring and Assessment of Adverse Drug Effects*. Geneva: CIOMS/WHO, 1986.
 42. Levy M, Miller D, Kaufman D, et al. Major upper gastrointestinal bleeding and the use of aspirin and other nonnarcotic analgesics. *Arch Intern Med* 1988; **148**: 281–5.
 43. Rosenberg L, Palmer JR, Kaufman DW, Strom BL, Schottenfeld D, Shapiro S. Breast cancer in relation to the occurrence and timing of induced and spontaneous abortion. *Am J Epidemiol* 1988; **127**: 981–9.
 44. Shapiro S. Case-control studies of acute and long-term adverse effects. In: *Adverse Drug Reactions. A Global Perspective*. Proceedings of the WHO Anniversary Symposium, Uppsala, 1988. Uppsala: Department of Drugs, National Board of Health and Welfare; 51–5.
 45. Miller DR, Rosenberg L, Kaufman DW, Stolley P, Warshauer ME, Shapiro S. Breast cancer before age 45 and oral contraceptive use: new findings. *Am J Epidemiol* 1989; **129**: 269–80.
 46. Kaufman DW, Palmer JR, Rosenberg L, Stolley P, Warshauer E, Shapiro S. Tar content of cigarettes in relation to lung cancer. *Am J Epidemiol* 1989; **129**: 703–11.
 47. Kaufman DW, Kelly JP, Rosenberg L, Stolley PD, Warshauer ME, Shapiro S. Hydralazine use in relation to cancers of the lung, colon, and rectum. *Eur J Clin Pharmacol* 1989; **36**: 259–64.
 48. Rosenberg L, Palmer JR, Shapiro S. Gestational trophoblastic disease and oral-contraceptive use. *Am J Obstet Gynecol* 1989; **161**: 1087–8.
 49. Rosenberg L, Werler MM, Palmer JR, et al. The risks of cancers of the colon and rectum in relation to coffee consumption. *Am J Epidemiol* 1989; **130**: 895–903.
 50. Palmer JR, Rosenberg L, Kaufman DW, Warshauer ME, Stolley P, Shapiro S. Oral contraceptive use and liver cancer. *Am J Epidemiol* 1989; **130**: 878–82.
 51. Kaufman DW, Kelly JP, Welch WR, et al. Noncontraceptive estrogen use and epithelial ovarian cancer. *Am J Epidemiol* 1989; **130**: 1142–51.
 52. Kelly JP, Rosenberg L, Kaufman DW, Shapiro S. Reliability of personal interview data in a hospital-based case-control study. *Am J Epidemiol* 1990; **131**: 79–90.
 53. Kaufman DW, Werler MM, Palmer JR, et al. Diazepam use in relation to breast cancer: results from two case-control studies. *Am J Epidemiol* 1990; **131**: 483–90.
 54. Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Stolley PD, Shapiro S. Vasectomy and the risk of prostate cancer. *Am J Epidemiol* 1990; **132**: 1051–5.
 55. Austin KL, Palmer JR, Seddon JM, et al. Case control study of idiopathic retinal detachment. *Int J Epidemiol* 1990; **19**: 1045–9.
 56. Miller DR. Breast cancer and early oral contraceptive use: studies from the Slone Epidemiology Unit. In: Mann RD, ed., *Oral Contraceptives and Breast Cancer*. Park Ridge, NJ: Parthenon, 1990; 137–57.
 57. Rosenberg L, Palmer JR, Shapiro S. Oral contraceptive use in relation to the risk of breast cancer: the data of Slone Epidemiology Unit in women aged 45–59 years. In: Mann RD, ed., *Oral Contraceptives and Breast Cancer*. Park Ridge, NJ: Parthenon, 1990; 159–68.
 58. Shapiro S, Rosenberg L. The need for future studies of breast cancer risk in relation to oral contraceptive use: a comprehensive strategy. In: Mann RD, ed., *Oral Contraceptives and Breast Cancer*. Park Ridge, NJ: Parthenon, 1990; 373–94.
 59. Palmer JR. Oral contraceptive use and gestational choriocarcinoma. *Cancer Detect Prev* 1990; **15**: 45–8.
 60. 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.
 61. Lesko SM, Rosenberg L, Kaufman DW, et al. Endometrial cancer and age at last delivery. Evidence of an association. *Am J Epidemiol* 1991; **133**: 554–9.

62. Palmer JR, Rosenberg L, Clarke EA, *et al.* Breast cancer and cigarette smoking: a hypothesis. *Am J Epidemiol* 1991; **134**: 1–13.
63. Johannes CB, Kaufman DW, Rosenberg L, *et al.* Side of origin of epithelial ovarian cancer. *Br Med J* 1992; **304**: 27–8.
64. Kaufman DW, Palmer JR, de Mouzon J, *et al.* Estrogen replacement therapy and the risk of breast cancer: results from the case-control surveillance study. *Am J Epidemiol* 1991; **134**: 1375–85.
65. Palmer JR, Rosenberg L, Strom BL, *et al.* Oral contraceptive use and risk of cutaneous malignant melanoma. *Cancer Causes Control* 1992; **3**: 547–54.
66. Kaufman DW, Palmer JR, Rosenberg L, Shapiro S. Oestrogen replacement therapy and breast cancer: new results from the Slone Epidemiology Unit's Case-Control Surveillance Study. In: Mann RD, ed., *Hormone Replacement Therapy and Breast Cancer Risk*. Park Ridge, NJ: Parthenon, 1992; 173–84.
67. Rosenberg L, Palmer JR, Zauber AG, *et al.* A case-control study of oral contraceptive use and invasive epithelial ovarian cancer. *Am J Epidemiol* 1994; **139**: 654–61.
68. Rosenberg L, Palmer JR, Zauber AG, *et al.* The relation of vasectomy to the risk of cancer. *Am J Epidemiol* 1994; **140**: 431–8.
69. Palmer JR, Rosenberg L, Harlap S, *et al.* Adult height and risk of breast cancer among U.S. black women. *Am J Epidemiol* 1995; **141**: 845–9.
70. Rosenberg L, Palmer JR, Zauber AG, *et al.* The relation of benzodiazepine use to the risk of selected cancers: breast, large bowel, malignant melanoma, lung, endometrium, ovary, non-Hodgkins lymphoma, testis, Hodgkin's disease, thyroid and liver. *Am J Epidemiol* 1995; **141**: 1153–60.
71. Palmer JR, Rosenberg L, Rao RS, *et al.* Oral contraceptive use and breast cancer risk among African-American women. *Cancer Causes Control* 1995; **6**: 321–31.
72. Rosenberg L, Palmer JR, Rao RS, *et al.* Case-control study of oral contraceptive use and risk of breast cancer. *Am J Epidemiol* 1996; **143**: 25–37.
73. Zhang Y, Rosenberg L, Colton T, *et al.* Adult height and risk of breast cancer among white women in a case-control study. *Am J Epidemiol* 1996; **143**: 1123–8.
74. Palmer JR, Rosenberg L, Rao RS, *et al.* Induced and spontaneous abortion in relation to risk of breast cancer (United States). *Cancer Causes Control* 1997; **8**: 841–9.
75. Rosenberg L, Rao RS, Palmer JP, *et al.* Transitional cell cancer of the urinary tract and renal cell cancer in relation to acetaminophen use. *Cancer Causes Control* 1998; **9**: 83–8.
76. Rosenberg L, Rao RS, Palmer JR, *et al.* Calcium channel blockers and the risk of cancer. *J Am Med Assoc* 1998; **279**: 1000–4.

REFERENCES

1. Doll R. Conclusions. In: Colombo F, Shapiro S, Slone D, Tognoni G, eds, *Epidemiological Evaluation of Drugs*. Amsterdam: Elsevier-North-Holland, 1977; 330.
2. Slone D. Personal communication. 1975.
3. Slone D, Jick H, Borda I, *et al.* Drug surveillance utilizing nurse monitors. *Lancet* 1966; **2**: 901–3.
4. Finney DJ. The design and logic of a monitor of drug use. *J Chron Dis* 1965; **18**: 77–98.
5. Thoburn R, Johnson JE, Cluff LE. Studies on the epidemiology of adverse drug reactions. III. Reactions in patients on a general medical service. *Bull Johns Hopkins Hosp* 1966; **119**: 299–368.
6. Lenz W. Thalidomide and congenital abnormalities. *Lancet* 1962; **1**: 45.
7. Inman WHW, Vessey MP. Investigation of deaths from pulmonary, coronary and cerebral thrombosis and embolism in women of child-bearing age. *Br Med J* 1968; **2**: 193–9.
8. Stadel BV. Oral contraceptives and cardiovascular disease. *New Engl J Med* 1981; **305**: 612–18.
9. Wright P. Untoward effects associated with practolol administration. Oculocutaneous syndrome. *Br Med J* 1975; **1**: 595–9.
10. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *New Engl J Med* 1971; **284**: 878–81.
11. Sex hormones (II). In: WHO International Agency for Research on Cancer, *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, Vol. 21. Lyon: IARC, 1979; 147–59.
12. Sartwell PE, Masi AT, Arthes FG, Greene GR, Smith HE. Thromboembolism and oral contraceptives: an epidemiologic case-control study. *Am J Epidemiol* 1969; **90**: 365–80.
13. Finney DJ. Statistical logic in the monitoring of reactions to therapeutic drugs. In: Inman WHW, ed., *Monitoring for Drug Safety*. Lancaster: MTP Press, 1980; 383–400.
14. Feinstein AR. The epidemiologic triad, the ablative risk ratio, and "retrospective" research. *Clin Pharmacol Ther* 1973; **14**: 291–306.
15. Hurwitz M, Wade OL. Intensive hospital monitoring of adverse reactions to drugs. *Br Med J* 1969; **1**: 531–6.
16. Friedman GD, Collen MF, Harris LE, Van Brunt EE, Davis LS. Experience in monitoring drug reactions in outpatients: the Kaiser-Permanente Drug Monitoring System. *J Am Med Assoc* 1971; **217**: 567–72.
17. Jick H, Miettinen OS, Shapiro S, *et al.* Comprehensive drug surveillance. *J Am Med Assoc* 1970; **213**: 1455–60.

18. Boston Collaborative Drug Surveillance Program. Regular aspirin intake and acute myocardial infarction. *Br Med J* 1974; **1**: 440–3.
19. Weiss HJ, Aledort LM, Kochwa S. The effect of salicylates on the hemostatic properties of platelets in man. *J Clin Invest* 1968; **47**: 2169–85.
20. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *New Engl J Med* 1989; **321**: 129–35.
21. Levy M. Aspirin use in patients with major upper gastrointestinal bleeding and peptic ulcer disease. *New Engl J Med* 1974; **290**: 1158–62.
22. Boston Collaborative Drug Surveillance Program. Oral contraceptives and venous thromboembolic disease, surgically confirmed gallbladder disease, and breast tumors. *Lancet* 1973; **1**: 1399–404.
23. Boston Collaborative Drug Surveillance Program. Functional ovarian cysts and oral contraceptives: negative association confirmed surgically. *J Am Med Assoc* 1974; **228**: 68–9.
24. Cancer and Steroid Hormone Study. The reduction in the risk of ovarian cancer associated with oral contraceptive use. *New Engl J Med* 1987; **316**: 650–5.
25. Rosenberg L, Shapiro S, Slone D, *et al.* Epithelial ovarian cancer and the use of combination oral contraceptives. *J Am Med Assoc* 1982; **247**: 3210–12.
26. Shapiro S, Kaufman DW, Slone D, *et al.* Recent and past use of conjugated estrogens in relation to adenocarcinoma of the endometrium. *New Engl J Med* 1980; **303**: 485–9.
27. Shapiro S, Parsells JL, Rosenberg L, *et al.* Risk of localized and widespread endometrial cancer in relation to recent and discontinued use of conjugated estrogens. *New Engl J Med* 1985; **313**: 969–72.
28. Boston Collaborative Drug Surveillance Program. Reserpine and breast cancer. *Lancet* 1974; **2**: 669–71.
29. Anonymous. The case-control study: consensus and controversy. *J Chron Dis* 1979; **32**: 1–144.
30. Anonymous. Some pharmaceutical drugs. In: WHO International Agency for Research on Cancer, *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*. Vol. 24. Lyon: IARC, 1980; 211–41.
31. Feinstein AR. Scientific standards in epidemiologic studies of the menace of daily life. *Science* 1988; **242**: 1257–64.
32. Kass EH, Shapiro S, Feinstein AR. Letters to the Editor. *Science* 1989; **243**: 1255–8.
33. Shapiro S, Parsells JL, Rosenberg L, Kaufman DW, Stolley PD, Schottenfeld D. Risk of breast cancer in relation to the use of rauwolfia alkaloids. *Eur J Clin Pharmacol* 1984; **26**: 143–6.
34. Helmrich SP, Rosenberg L, Kaufman DW, *et al.* Lack of an elevated risk of malignant melanoma in relation to oral contraceptive use. *J Natl Cancer Inst* 1984; **72**: 617–20.
35. Palmer JR, Rosenberg L, Strom BL, *et al.* Oral contraceptive use and risk of cutaneous malignant melanoma. *Cancer Causes Control* 1992; **3**: 547–54.
36. Levy M, Miller D, Kaufman D, *et al.* Major upper gastrointestinal bleeding and the use of aspirin and other nonnarcotic analgesics. *Arch Intern Med* 1988; **148**: 281–5.
37. 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.
38. Rothman KJ. *Modern Epidemiology*. Boston: Little Brown, 1986.
39. Poulsen HE, Loft S. The impact of genetic polymorphisms in risk assessment of drugs [review]. *Arch Toxicol* 1994; **16** (suppl.): 211–22.
40. Richards B, Skoletsky J, Shuber AP, *et al.* Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal swabs/brushes. *Hum Mol Genet* 1993; **2**: 159–63.
41. Rosenberg L, Slone D, Shapiro S, *et al.* Breast cancer and alcoholic-beverage consumption. *Lancet* 1982; **1**: 267–70.
42. Rosenberg L, Metzger LS, Palmer JR. Epidemiology of breast cancer: alcohol consumption. *Epidemiol Rev* 1993; **15**: 133–44.
43. Rosenberg L, Palmer JR, Miller DR, Clarke EA, Shapiro S. A case-control study of alcoholic beverage consumption and breast cancer. *Am J Epidemiol* 1990; **131**: 6–14.
44. Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Stolley PD, Shapiro S. Vasectomy and the risk of prostate cancer. *Am J Epidemiol* 1990; **132**: 1051–5.
45. Honda GD, Bernstein L, Ross RK, *et al.* Vasectomy, cigarette smoking and age at first sexual intercourse as risk factors for prostatic cancer in middle-aged men. *Br J Cancer* 1988; **57**: 326–31.
46. Guess HA. Invited commentary: vasectomy and prostate cancer. *Am J Epidemiol* 1990; **132**: 1062–3.
47. Weiss NS, Lyon JL, Krishnamurthy S, *et al.* Non-contraceptive estrogen use and the occurrence of ovarian cancer. *J Natl Cancer Inst* 1982; **68**: 95–8.
48. Kaufman DW, Kelly JP, Welch WR, *et al.* Non-contraceptive estrogen use and epithelial ovarian cancer. *Am J Epidemiol* 1989; **130**: 1142–51.
49. Palmer JR, Rosenberg L, Kaufman DW, Warshauer ME, Stolley P, Shapiro S. Oral contraceptive use and liver cancer. *Am J Epidemiol* 1989; **130**: 878–82.
50. Helmrich SP, Rosenberg L, Kaufman DW, Strom B, Shapiro S. Venous thromboembolism in relation to oral contraceptive use. *Obstet Gynecol* 1987; **69**: 91–5.
51. Shapiro S, Rosenberg L. The need for future studies of breast cancer risk in relation to oral contraceptive

- use: a comprehensive strategy. In: Mann RD, ed., *Oral Contraceptives and Breast Cancer*. Park Ridge, NJ: Parthenon, 1990; 373-94.
52. Maclure M, Greenland S. Tests for trend and dose response: misinterpretations and alternatives. *Am J Epidemiol* 1992; **135**: 96-104.
 53. Kaufman DW, Kelly JP, Levy M, Shapiro S. *The Drug Etiology of Agranulocytosis and Aplastic Anemia: the International Agranulocytosis and Aplastic Anemia Study*. New York: Oxford University Press, 1991.
 54. WHO International Agency for Research on Cancer. Supplement 7. Overall evaluations of carcinogenicity: an updating of IARC monographs from volumes 1-24. In: *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, Supplement 7. Lyon: IARC, 1987; 96-9.
 55. Surgeon General of the United States. *The Health Consequences of Involuntary Smoking*. Washington: USDHHS, Government Printing Office, 1986.
 56. Harlow SD, Linet MS. Agreement between questionnaire data and medical records. The evidence for accuracy of recall. *Am J Epidemiol* 1989; **129**: 233-48.
 57. Anonymous. Some miscellaneous pharmaceutical substances. Metronidazole. In: WHO International Agency for Research on Cancer, *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, Vol. 13. Lyon: IARC, 1977; 113-22.
 58. Strom BL. Other approaches to pharmacoepidemiologic studies. In: Strom BL, ed., *Pharmacoepidemiology*. New York: Churchill Livingstone, 1989; 235-6.
 59. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol* 1992; **135**: 1019-28.
 60. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol* 1992; **135**: 1029-41.
 61. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. III. Design options. *Am J Epidemiol* 1992; **135**: 1042-50.
 62. Miettinen OS, ed. Valid sampling in the context of a base secondary to the cases. In: *Theoretical Epidemiology. Principles of Occurrence Research in Medicine*. New York: Wiley, 1985; section 4.4.2, 79-83.
 63. Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. A policy perspective. *Ann Intern Med* 1979; **90**: 257-61.
 64. Spitzer WO, Suissa S, Ernst P, et al. The use of beta-agonists and the risk of death and near death from asthma. *New Engl J Med* 1992; **326**: 501-6.
 65. Pahor M, Guralnik JM, Salive ME, Corti M-C, Carbonin P, Havlik RJ. Do calcium channel blockers increase the risk of cancer? *Am J Hypertens* 1996; **9**: 695-9.
 66. Pahor M, Guralnik JM, Corti M-C, et al. Calcium-channel blockade and incidence of cancer in aged populations. *Lancet* 1996; **348**: 493-7.
 67. Rosenberg L, Rao RS, Palmer JR, et al. Calcium channel blockers and the risk of cancer. *J Am Med Assoc* 1998; **279**: 1000-4.
 68. Derby LE, Jick H. Acetaminophen and renal and bladder cancer. *Epidemiology* 1996; **7**: 358-62.
 69. Rosenberg L, Rao RS, Palmer JP, et al. Transitional cell cancer of the urinary tract and renal cell cancer in relation to acetaminophen use. *Cancer Causes Control* 1998; **9**: 83-8.
 70. Rosenberg L, Shapiro S, Slone D, Kaufman DW, Miettinen OS, Stolley PD. Thiazides and acute cholecystitis. *New Engl J Med* 1980; **303**: 546-8.
 71. Palmer JR. Oral contraceptive use and gestational choriocarcinoma. *Cancer Detect Prev* 1990; **15**: 45-8.
 72. Kakar F, Weiss NS, Strite SA. Thiazide use and the risk of cholecystectomy in women. *Am J Epidemiol* 1986; **124**: 428-33.
 73. Thun MJ, Mamboodiri MM, Heath CW. Aspirin use and reduced risk of fatal colon cancer. *New Engl J Med* 1991; **325**: 1593-6.
 74. Rosenberg L, Louik C, Shapiro S. Nonsteroidal anti-inflammatory drug use and reduced risk of large bowel carcinoma. *Cancer* 1998; **82**: 2326-33.
 75. Kaufman DW, Kelly JP, Rosenberg L, Stolley PD, Warshauer ME, Shapiro S. Hydralazine use in relation to cancers of the lung, colon and rectum. *Eur J Clin Pharmacol* 1989; **36**: 259-64.
 76. Kaufman DW, Shapiro S, Slone D, et al. Diazepam and the risk of breast cancer. *Lancet* 1982; **1**: 537-9.
 77. Horwitz RI, Feinstein AR. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. *New Engl J Med* 1978; **299**: 1089-94.
 78. Kaufman DW, Shapiro S, Rosenberg L, et al. Intrauterine contraceptive device use and pelvic inflammatory disease. *Am J Obstet Gynecol* 1980; **136**: 159-62.
 79. Kaufman DW, Watson J, Rosenberg L, et al. The effect of different types of intrauterine device upon the risk of pelvic inflammatory disease. *J Am Med Assoc* 1983; **250**: 759-62.
 80. Kaufman DW, Shapiro S, Slone D, et al. Decreased risk of endometrial cancer among oral contraceptive users. *New Engl J Med* 1980; **303**: 1045-7.
 81. Hurwitz ES, Barrett MJ, Bregman D, et al. Public Health Service study on Reye's syndrome and medication. Report of the pilot phase. *New Engl J Med* 1985; **313**: 849-57.
 82. Wiholm BE. Personal communication, 1991.

83. The International Agranulocytosis and Aplastic Anemia Study. Anti-infective drug use in relation to the risk of agranulocytosis and aplastic anemia. *Arch Intern Med* 1989; **149**: 1036–40.
84. Kaufman DW, Kelly JP, Johannes CB, *et al.* Acute thrombocytopenic purpura in relation to the use of drugs. *Blood* 1993; **82**: 2714–18.
85. Miettinen O, Slone D, Shapiro S. Current problems in drug-related epidemiologic research. In: Colombo F, Shapiro S, Slone D, Tognoni G, eds, *Epidemiological Evaluation of Drugs*. Amsterdam: Elsevier–North-Holland, 1977; 295–303.
86. Shapiro S. The role of automated record linkage in the postmarketing surveillance of drug safety. A critique. *Clin Pharmacol Ther* 1989; **46**: 371–86.
87. Shapiro S. Automated record linkage: a response to the commentaries and letters to the editor. *Clin Pharmacol Ther* 1989; **46**: 395–8.
88. Faich GA, Stadel B. The future of automated record linkage for postmarketing surveillance. *Clin Pharmacol Ther* 1989; **46**: 387–9.
89. Strom BL, Carson JL. Automated databases used for pharmacoepidemiology research. *Clin Pharmacol Ther* 1989; **46**: 390–4.
90. Jick H, Walker HM. Uninformed criticism of automated record linkage [letter to the editor]. *Clin Pharmacol Ther* 1989; **46**: 478–9.
91. Faich GA. Record linkage for postmarketing surveillance [letter to the editor]. *Clin Pharmacol Ther* 1989; **46**: 479–80.
92. Tilson H. Pharmacoepidemiology: the lessons learned; the challenges ahead [letter to the editor]. *Clin Pharmacol Ther* 1979; **46**: 480.
93. Sacks RM, Dieck GS. The feldene experience: a case study. In: Horisberger B, Dinkel R, eds, *The Perception and Management of Drug Safety Risks*. New York: Springer, 1989; 81–9.
94. Laporte J-R, Carne X, Vidal X, *et al.* Upper gastrointestinal bleeding in relation to previous use of analgesics and non-steroidal anti-inflammatory drugs. *Lancet* 1991; **337**: 85–9.
95. Kaufman DW, Kelly JP, Sheehan JE, Laszlo A, Wilhom BE, Alfredsson L. Nonsteroidal anti-inflammatory drug use in relation to major upper gastrointestinal bleeding. *Clin Pharmacol Ther* 1993; **53**: 485–94.

14

Prescription-event Monitoring

RONALD D. MANN

Drug Safety Research Unit, Southampton, UK

INTRODUCTION

On 16 December 1961 the *Lancet* published the first report¹ in the British medical press indicating that thalidomide might be a human teratogen. Following the thalidomide tragedy, a Committee on Safety of Drugs was established by the Health Ministers in the UK to deal, on a voluntary basis, with the problem of drug registration. This committee, at first chaired by Sir Derrick Dunlop, operated while legislation could be enacted. Its final report, for 1969 and 1970, preceded the implementation of the Medicines Act of 1968, and contained the following remarkable paragraph:²

no drug which is pharmacologically effective is entirely without hazard. The hazard may be insignificant or may be acceptable in relation to the drug's therapeutic action. Furthermore, not all hazards can be known before a drug is marketed; neither tests in animals nor clinical trials in patients will always reveal all the possible side effects of a drug. These may only be known when the drug has been administered to large numbers of patients over considerable periods of time.

Thus, for almost 30 years it has been well recognized that drug safety depends upon post-marketing surveillance.

The essential reason for this is that the clinical safety database for new drugs is, in epidemiological terms, restricted both in size and in patient characteristics. In the UK, for example, successful applications for product licenses for medicines containing new active substances include safety information on a median of only 1480 (range 129–9400) patients.³ Most of these patients will have had only one well identified disease treated with only one drug. Few, if any of them, will be typical of the elderly patients with more than one disease who are likely to receive several drugs in everyday clinical practice.

The safety database at the time of licensing may, therefore, be only very poorly generalizable to the population that may be exposed to the drug once it is marketed.

For these and related reasons, at least 13 clinical safety withdrawals of new drugs licensed in the UK occurred in the 20 years following the implementation of the Medicines Act of 1968.⁴ It became clear that the modern drug regulatory apparatus, including the whole mechanism of spontaneous reporting that had been established by the Committee on Safety of Drugs in 1964⁵ required some form of augmentation.

The Committee on Safety of Medicines (the successor to the Committee on Safety of Drugs),

wishing to consider these matters, established a working party under the chairmanship of Professor David Grahame-Smith. The working party 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”.⁶ It is these proposals that form the purpose and basis of what has now become known as prescription-event monitoring (PEM)—a non-interventional observational cohort technique first established in England by Dr William Inman, which remains a principal activity of the Drug Safety Research Unit at Southampton. PEM is one form of pharmacovigilance that, with the development and harmonization of drug regulation in the European Community, has its legal basis in Directives 65/65 and 75/319, and in Regulation 2309/93.

DESCRIPTION

PEM is noninterventional in the sense that there is no interference with the doctor’s decision regarding what to prescribe 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 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 GP issues prescriptions for the medicines considered medically necessary. The patient takes the prescription to a pharmacist who dispenses the medication and then sends the prescription to a central Prescription Pricing Authority (PPA) which effects reimbursement. The Drug Safety Research Unit (DSRU) is, under long standing and confidential arrangements, provided with electronic copies of all those

prescriptions issued throughout England for the drugs being monitored. These arrangements operate for the length of time necessary for the DSRU to collect the first 50 000 prescriptions that identify 20–30 000 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 prescribed and dispensed, but there is no measure of compliance. Information is available on coprescribed drugs but not on the use of nonprescription medication. Collection of the exposure data begins immediately after the new drug has been launched. The drugs to be monitored are chosen by the DSRU, preference being given to medicines intended for widespread, long term use.

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, but no GP is sent more than four green forms in any one month.

The green form is illustrated in Figure 14.1, which shows the definition of an “event” and the other information requested of the GP. The GP is not paid to provide this information, which is provided, under conditions of medical confidence, in the interests of drug safety. The system provides good contact with the GPs and facilitates the collection of any followup or additional data considered necessary by the research physicians monitoring each study and working within the DSRU. All pregnancies reported during treatment or within three months of stopping the drug are followed up to determine the outcome, and any deaths for which the cause is not known or which may have been related to medication are followed up by contact with the GP, the Office for National Statistics, or the health authorities of the National Health Service.

PLEASE REMOVE THIS SECTION OF THE FORM SO THAT THE BOTTOM HALF BECOMES ANONYMOUS.

<p>DRUG SAFETY RESEARCH UNIT PRESCRIPTION EVENT MONITORING</p> <p>MEDICAL - IN CONFIDENCE</p>	<p>Saad A. W. Shakir, FRCP (Glas & Ed), FFPM, MRCGP Bursledon Hall, Southampton SO31 1AA Telephone: (023) 80408600</p> <p>We collect EVENT data.</p> <p>An EVENT is any new diagnosis, any reason for referral to a consultant or admission to hospital, any unexpected deterioration (or improvement) in a concurrent illness, any suspected drug reaction, any alteration of clinical importance in laboratory values or any other complaint which was considered of sufficient importance to enter in the patient's notes. Example: A broken leg is an EVENT.</p> <p>Please indicate if you suspect an EVENT to be an adverse reaction to a drug.</p> <p>These studies are conducted in accordance with the results of authoritative discussions and the International Ethical Guidelines for Biomedical Research Involving Human Subjects prepared by the Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization (WHO), Geneva 1993. The method of study also complies with the Guidelines on the practice of Ethics Committees in Medical Research involving Human Subjects, as issued by the Royal College of Physicians of London (August 1996)</p> <p>Ref:</p>
--	--

The DSRU is advised by the Drug Safety Research Trust, a registered independent charity (No. 327206) The unit operates in association with the University of Southampton. Trustees Professor C. G. George MD FRCP, Sir Gordon Higginson DL FEng, Professor Stephen T Holgate MD DSc FRCP, Professor M. P. Vessey CBF MD FRCP FFPHM FRS.

PLEASE RETURN THIS HALF OF THE FORM. Ref:

Sex	Date of Birth / /	Was the drug effective?	Yes <input type="checkbox"/> No <input type="checkbox"/> Don't know <input type="checkbox"/>	
Indication for prescribing		Has drug been stopped?	Yes <input type="checkbox"/> No <input type="checkbox"/> Don't know <input type="checkbox"/>	
Start date ___/___/___	Dose ___ mg/day	If 'Yes' reason for stopping		
Date of last prescription ___/___/___		Date of stopping if known ___/___/___		
Please list any asthma medication (including steroids) that this patient has taken since starting this drug (please continue overleaf if necessary)				
Event Date	Dose daily	Events while taking this drug If none, please tick box <input type="checkbox"/>	Event Date	Events after stopping this drug If none, please tick box <input type="checkbox"/>

IMPORTANT: PLEASE INDICATE ANY EVENT REPORTED TO CSM OR MANUFACTURER

Figure 14.1.

Table 14.1. Cohort sizes for 65 drugs

	Generic name	Drug name	Group	GF sent	GF returned	Voids	Cohort	Male	Female
1	Cisapride	Prepulsid	Antispasmodic	23 133	14 442	1 208	13 234	5 485	7 623
2	Famotidine	Pepcid	H ₂ -antagonist	20 399	10 560	1 060	9 500	4 899	4 396
3	Nizatidine	Axid	H ₂ -antagonist	19 454	8 691	909	7 782	4 098	3 555
4	Misoprostol	Cytotec	Prostaglandin analogue	22 034	14 836	1 061	13 775	4 939	8 592
5	Lansoprazole	Zoton	Proton pump inhibitor	36 722	18 742	1 413	17 329	8 160	8 975
6	Omeprazole	Losec	Proton pump inhibitor	28 496	17 772	1 568	16 204	7 968	8 073
7	Pantoprazole	Prorium	Proton pump inhibitor	28 159	12 521	980	11 541	5 350	6 068
8	Betaxolol	Kerlone	β -blocker	3 187	1 743	212	1 531	644	852
9	Doxazosin	Cardura	α -blocker	15 498	9 315	833	8 482	3 799	4 622
10	Enalapril	Innovace	ACE inhibitor	24 719	16 875	1 514	15 361	7 081	7 951
11	Lisinopril	Zestril + Carace	ACE inhibitor	20 631	13 091	653	12 438	5 469	6 712
12	Perindopril	Coversyl	ACE inhibitor	18 760	10 026	937	9 089	4 094	4 935
13	Ramipril	Tritace	ACE inhibitor	3 287	1 556	185	1 371	618	739
14	Losartan	Cozaar	Antihypertensive	26 303	15 755	1 233	14 522	5 834	8 617
15	Amlodipine	Istin	Ca antagonist	24 140	14 177	1 208	12 969	6 085	6 751
16	Diltiazem	Tildiem	Ca antagonist	16 665	11 219	1 107	10 112	6 000	3 972
17	Isradipine	Prescal	Ca antagonist	8 073	4 139	460	3 679	1 515	2 128
18	Nicardipine	Cardene	Ca antagonist	19 243	12 048	1 138	10 910	5 276	5 484
19	Nicorandil	Ikorel	K-channel activator	25 451	14 826	1 206	13 620	8 323	5 219
20	Xamoterol	Corwin	Inotropic	8 561	5 882	509	5 373	2 846	2 467
21	Fluvastatin	Lescol	Lipid lowering	12 883	8 143	601	7 542	4 075	3 427
22	Bambuterol	Bambec	β_2 -agonist	18 013	9 149	1 051	8 098	3 631	4 400
23	Salmeterol	Serevent	β_2 -agonist	28 019	17 347	1 940	15 407	7 844	7 445
24	Nedocromil	Tilade	Asthma prophylaxis	20 644	14 057	1 763	12 294	6 340	5 768
25	Acrivastine	Semprex	Antihistamine	15 772	8 911	1 048	7 863	2 833	4 899
26	Cetirizine	Zirtek	Antihistamine	18 933	10 877	1 323	9 554	3 945	5 457
27	Fexofenadine	Telfast	Antihistamine	35 817	18 238	1 600	16 638	6 578	9 880
28	Loratadine	Claritin	Antihistamine	21 084	10 684	1 376	9 308	3 912	5 179
29	Flunitrazepam	Rohypnol	Benzodiazepine	NK	9 325	1 833	7 492	2 368	4 951
30	Zolpidem	Stilnoct	Hypnotic	30 986	15 169	1 709	13 460	4 986	8 305
31	Zopiclone	Zimovane	Hypnotic	24 026	13 177	1 634	11 543	3 989	7 461
32	Buspirone	Buspar	Anxiolytic	24 185	13 087	1 974	11 113	3 500	7 419

33	Risperidone	Risperdal	Antipsychotic	14 189	9 174	1 490	7 684	4 124	3 500
34	Moclobemide	Manerix	MAOI	21 287	12 518	1 683	10 835	3 940	6 824
35	Fluoxetine	Prozac	SSRI	24 738	14 444	1 752	12 692	3 690	8 863
36	Fluoxamine	Faverin	SSRI	20 504	12 279	1 296	10 983	3 094	7 694
37	Paroxetine	Seroxat	SSRI	25 843	15 907	2 166	13 741	4 373	9 279
38	Sertraline	Lustral	SSRI	24 632	14 817	2 083	12 734	3 910	8 729
39	Nefazodone	Dutonin	Antidepressant	24 351	13 365	1 531	11 834	4 418	7 347
40	Venlafaxine	Efexor	Antidepressant	25 870	14 119	1 477	12 642	4 349	8 214
41	Tramadol	Zydol	Analgesic	20 546	11 470	938	10 532	3 869	6 499
42	Sumatriptan	Imigran	Antimigraine	23 625	16 724	1 796	14 928	2 881	11 948
43	Lamotrigine	Lamictal	Anti-epileptic	19 448	13 196	1 880	11 316	5 646	5 637
44	Vigabatrin	Sabril	Anti-epileptic	17 000	11 769	1 591	10 178	5 150	5 023
45	Gabapentin	Neurontin	Anti-epileptic	5 285	3 508	408	3 100	1 501	1 587
46	Cefixime	Suprax	Cephalosporin	32 526	12 880	1 630	11 250	4 799	6 223
47	Azithromycin	Zithromax	Macrolide	23 900	12 535	1 260	11 275	4 532	6 575
48	Ciprofloxacin	Ciproxin	Quinolone	20 664	12 394	917	11 477	4 493	6 612
49	Enoxacin	Comprecin	Quinolone	7 369	3 281	491	2 790	475	2 276
50	Norfloxacin	Utinor	Quinolone	26 036	13 029	1 919	11 110	1 852	9 098
51	Ofloxacin	Tarivid	Quinolone	27 787	12 698	1 665	11 033	4 263	6 629
52	Fosfomycin	Monuril	Antibacterial	8 303	3 783	420	3 363	286	3 033
53	Fluconazole	Diflucan	Antifungal	23 846	16 347	1 332	15 015	877	14 017
54	Itraconazole	Sporanox	Antifungal	24 216	15 370	1 725	13 645	1 482	12 102
55	Aciclovir	Zovirax	Antiviral	17 027	12 611	1 560	11 051	3 953	6 866
56	Famciclovir	Famvir	Antiviral	23 420	15 311	1 142	14 169	5 578	8 449
57	Valaciclovir	Valtrex	Antiviral	21 725	13 926	1 122	12 804	4 935	7 775
58	Acarbose	Glucobay	Antidiabetic	23 910	15 004	1 349	13 655	6 442	7 074
59	Finasteride	Proscar	Prostate treatment	25 562	16 093	1 321	14 772	14 767	2
60	Alendronate	Fosamax	Biphosphonate	22 131	13 135	1 219	11 916	1 240	10 619
61	Terodiline	Terolin	Anticholinergic	20 069	13 971	1 527	12 444	3 378	8 912
62	Etodolac	Lodine	NSAID	20 725	10 334	1 243	9 091	3 002	5 925
63	Meloxicam	Mobic	NSAID	39 147	20 360	1 274	19 087	6 174	12 588
64	Nabumetone	Relifex	NSAID	21 007	11 540	1 096	10 444	3 437	6 838
65	Tenoxicam	Mobiflex	NSAID	26 357	11 731	849	10 882	3 702	6 940
			Total	1 366 322	796 003	82 398	713 606	283 066	422 019
			Mean		12 246		10 979		
			Response rate		57.6	10.4	51.9		

Table 14.2. Collection periods for 65 drugs

	Generic name	Group	1st Rx	Last Rx used	Collection time (mths) for Rx	Rx on database	GF send interval approx	Last GF sent	Total time (mths)
1	Cisapride	Antispasmodic	Oct 90	Apr 91	7	38 291	6	May 92	20
2	Famotidine	H ₂ -antagonist	Nov 87	Mar 88	5	31 809	6	Sep 88	12
3	Nizatidine	H ₂ -antagonist	Sep 87	Sep 88	13	35 474	12	Oct 89	26
4	Misoprostol	Prostaglandin analogue	Oct 88	Jul 89	10	38 624	6	May 90	20
5	Lansoprazole	Proton pump inhibitor	May 94	Nov 94	7	57 216	6	Nov 95	19
6	Omeprazole	Proton pump inhibitor	Jun 89	Jun 90	13	58 562	12	Oct 91	28
7	Pantoprazole	Proton pump inhibitor	Dec 96	Jun 97	7	42 135	6	Apr 98	17
8	Betaxolol	β -blocker	Nov 84	Jun 87	32	16 527	12	Jun 88	44
9	Doxazosin	α -blocker	Mar 89	Jan 91	23	72 996	6	Mar 92	37
10	Enalapril	ACE inhibitor	Apr 85	Jan 86	10	75 368	12	Jan 87	22
11	Lisinopril	ACE inhibitor	Jun 88	Mar 89	10	59 632	6	Sep 89	16
12	Perindopril	ACE inhibitor	Mar 90	Sep 93	43	95 164	12	Dec 94	58
13	Ramipril	ACE inhibitor	Apr 90	Feb 91	11	9 276	12	May 92	26
14	Losartan	Antihypertensive	May 95	Oct 96	18	81 632	6	May 97	25
15	Amlodipine	Ca antagonist	Mar 90	Mar 91	13	80 518	6	Jun 92	28
16	Diltiazem	Ca antagonist	Sep 84	Oct 86	26	63 620	12	Oct 87	38
17	Isradipine	Ca antagonist	Mar 89	Feb 91	24	38 792	6	Mar 92	37
18	Nicardipine	Ca antagonist	Nov 86	May 88	19	82 222	12	Jul 89	33
19	Nicorandil	K-channel activator	Dec 94	Oct 96	23	86 165	6	May 97	30
20	Xamoterol	Inotropic	Aug 88	Nov 90	28	49 803	6	Apr 91	33
21	Fluvastatin	Lipid lowering	Feb 94	Feb 96	25	44 342	6	Sep 96	31
22	Bambuterol	β_2 -agonist	Feb 93	Dec 95	35	61 368	6	Jul 96	42
23	Salmeterol	β_2 -agonist	Dec 90	May 91	6	42 895	12	Dec 92	25
24	Nedocromil	Asthma prophylaxis	Nov 86	Sep 88	23	60 405	12	Oct 89	36
25	Acrivastine	Antihistamine	May 89	Sep 90	17	21 794	6	Jan 91	21
26	Cetirizine	Antihistamine	May 89	Jul 89	3	20 789	6	Aug 90	16
27	Fexofenadine	Antihistamine	Mar 97	Aug 97	6	55 446	3	Jan 98	11
28	Loratadine	Antihistamine	May 89	Aug 89	4	23 173	6	Aug 90	16
29	Flunitrazepam	Benzodiazepine	Oct 82	Nov 83	14	0	12	Nov 84	26
30	Zolpidem	Hypnotic	Jul 94	Jan 96	19	51 681	6	Sep 96	27
31	Zopiclone	Hypnotic	Mar 91	Jul 91	5	45 627	6	Sep 92	19

32	Bupirone	Anxiolytic	Mar 88	Feb 89	12	40 950	12	40 950	Apr 90	26
33	Risperidone	Antipsychotic	Jul 93	Apr 96	33	67 977	6	67 977	Sep 96	38
34	Moclobemide	MAOI	Jun 93	Aug 95	27	66 752	6	66 752	Apr 96	35
35	Fluoxetine	SSRI	Mar 89	Mar 90	13	52 607	6	52 607	Mar 91	25
36	Fluoxamine	SSRI	Feb 87	Feb 88	13	34 592	12	34 592	Mar 89	26
37	Paroxetine	SSRI	Mar 91	Mar 92	13	48 508	6	48 508	Feb 93	24
38	Sertraline	SSRI	Jan 91	Sep 92	21	50 664	6	50 664	Aug 93	32
39	Nefazodone	Antidepressant	Jan 96	Feb 97	14	51 158	6	51 158	Oct 97	22
40	Venlafaxine	Antidepressant	May 95	Jun 96	14	54 549	6	54 549	Feb 97	22
41	Tramadol	Analgic	Jul 94	Dec 94	6	73 179	6	73 179	Jul 95	13
42	Sumatriptan	Antimigraine	Nov 91	Jul 92	9	58 625	6	58 625	May 93	19
43	Lamotrigine	Anti-epileptic	Dec 91	Feb 95	39	155 581	6	155 581	Sep 95	46
44	Vigabatrin	Anti-epileptic	Dec 91	Nov 94	36	152 200	6	152 200	Jun 95	43
45	Gabapentin	Anti-epileptic	Aug 93	Apr 95	21	28 432	6	28 432	Oct 95	27
46	Cefixime	Cephalosporin	Sep 90	May 91	9	35 426	6	35 426	Jul 92	23
47	Azithromycin	Macrolide	Mar 92	Jun 93	16	30 389	6	30 389	Feb 94	24
48	Ciprofloxacin	Quinolone	Nov 88	Jan 89	3	22 776	6	22 776	Jul 89	9
49	Enoxacin	Quinolone	Apr 89	Jan 91	22	8 658	12	8 658	Oct 91	31
50	Norflaxacin	Quinolone	Oct 90	Oct 91	13	28 443	12	28 443	Feb 93	29
51	Ofloxacin	Quinolone	May 90	Dec 91	20	30 249	12	30 249	Feb 93	34
52	Fosfomycin	Antibacterial	Feb 94	Jun 96	29	9 558	6	9 558	Oct 96	33
53	Fluconazole	Antifungal	Sep 88	Jan 89	5	25 546	6	25 546	Jun 89	10
54	Itraconazole	Antifungal	Apr 89	Apr 90	13	28 228	6	28 228	Jan 91	22
55	Aciclovir	Antiviral	Nov 85	Dec 86	14	20 224	6	20 224	May 87	19
56	Famciclovir	Antiviral	Feb 94	Aug 95	19	24 484	6	24 484	Apr 96	27
57	Valaciclovir	Antiviral	Mar 95	Feb 97	24	22 851	6	22 851	Sep 97	31
58	Acarbose	Antidiabetic	Jul 93	Aug 95	26	98 171	6	98 171	Mar 96	33
59	Finasteride	Prostate treatment	Oct 92	Feb 94	17	81 177	12	81 177	Mar 95	29
60	Alendronate	Biphosphonate	Oct 95	Jan 97	16	67 422	12	67 422	Mar 98	30
61	Terodiline	Anticholinergic	Nov 86	Sep 87	11	43 041	12	43 041	Dec 88	26
62	Etodolac	NSAID	Nov 85	Aug 86	10	30 087	12	30 087	Nov 87	25
63	Meloxicam	NSAID	Dec 96	Mar 97	4	65 805	5	65 805	Nov 97	12
64	Nabumetone	NSAID	Mar 87	Sep 87	7	30 794	12	30 794	Nov 88	21
65	Tenoxicam	NSAID	Nov 88	Aug 89	10	44 161	6	44 161	Mar 90	17
	Total				1 058	3 224 646		3 224 646		
	Mean				16.28	50 385		50 385		

Thus, in PEM the exposure data are derived from the NHS prescriptions for the drug being monitored, and the outcome data are obtained from the green form questionnaires and the followup process.

PEM collects event data and does not ask the doctor to determine whether any particular event represents an adverse drug reaction (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.

PEM studies aim to collect clinical information on some 10 000 patients and this is routinely achieved unless the drug fails to establish itself in the market.

Each green form is seen on the day of receipt by the medical or scientific officer monitoring the study in the DSRU. This initial review aims to identify serious possible ADRs or events requiring immediate followup. It uses the previously published⁷ CSM (Committee on Safety of Medicines) Serious Adverse Drug Reactions List.

Interim analyses of the computerized data are undertaken every 2500 patients in each study and contacts are, whenever possible, maintained with the company holding the product license, so that they (although the study is independent of them) can comply with the ADR/drug safety reporting procedures of the regulatory authorities.

Further details of the methods of data coding, DSRU computerized event reporting dictionary, and methods of computerization and analysis have been given in recent publications.⁸⁻¹¹ The DSRU is an independent registered medical charity (No. 327206), but is extensively supported by donations and grants from the pharmaceutical industry.

As shown in Table 14.1a, 65 PEM studies have been conducted to date, with a mean response rate (percentage of green forms returned) of $57.94 \pm 7.95\%$. The mean cohort size has been 10 979 patients, after allowing for forms which contain information but not event data (voids). The collection periods (the time for which is has been necessary to collect prescriptions yielding a finished cohort size averaging over 10 000 patients) vary markedly. As shown in Tables 1 and 2, the meloxicam cohort was of 19 087 patients and with this (very successful) drug, it was necessary to collect prescriptions for only four months. At the

opposite extreme, it was necessary to collect prescriptions for perindopril for 43 months, in order to achieve a finished cohort size of 9089 patients. Over all 65 studies, the mean collection period has been 16.28 (median 14; interquartile range 10 and 23) months.

STRENGTHS

PEM has a number of important strengths. First, as indicated above, the method is noninterventional in nature and does not interfere with the treatment the doctor chooses and advises as most appropriate for the individual patient. In this way, the technique avoids the selection biases involved in ordinary 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.

Third, 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 accrue from the early proposals of Professor DJ Finney²⁸ on event reporting.

Fourth, the method allows close contact between the research staff in the DSRU and the reporting doctors. This facilitates followup of important events, pregnancies, deaths, etc., and allows for the maximum clinical understanding of confounders, biases, and similar issues.

Fifth, 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 and is probably the reason why ADR reporting is more complete in PEM than in spontaneous ADR reporting systems, such as the yellow card system in the UK (see Chapters 10 and 11).

Sixth, 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 market place, 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.

Finally, the method provides a technique that can generate signals or hypotheses which can themselves be refuted or confirmed by the established methods of epidemiology.

WEAKNESSES

Of course, PEM also has important weaknesses. Not all of the green forms are returned and this could conceal undisclosed sources of bias. Studies designed to show whether patients whose doctors do not return the green forms are different from those of responding doctors are currently in progress.

Second, PEM is currently restricted to general practice, although pilot studies in hospital are currently under way. The pilot studies are progressing satisfactorily and will be reported shortly. It is planned to develop a regional system if the pilot studies prove satisfactory.

Third, there is no measure of compliance using dispensed prescriptions, i.e., it is not known whether the patient actually took the dispensed medication.

Fourth, the outcome data and medical followup are collected by nonelectronic, labor-intensive methods, and this disadvantage is only gradually being overcome as more and more doctors run their practices totally by computer.

Finally, issues of patient privacy and confidentiality are sensitive and rely heavily on doctor-to-doctor contact pursued in the public interest. This question of confidentiality currently occupies much attention, both in the US and throughout Europe (see Chapter 26).

PARTICULAR APPLICATIONS

SEARCHING FOR SIGNALS

Horrors List

The initial search is provided by the daily inspection of newly received green forms and their

examination for adverse events of a typically iatrogenic kind, such as many of those given on the previously mentioned list.⁷ The green form questionnaire asks the doctor to record the reason why the drug was withdrawn if, in fact, it was withdrawn and it has been found that this is a most informative question.

Display of Events During the Study/Events on Drug

Table 14.3 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 1900 or so events on the DSRU dictionary. The number of events reported is shown for each month of the study. Table 14.4 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 (and 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 65 drugs that have been studied, or by comparison with drugs of the same therapeutic group or with the same indication for use. The trend of reports over the months of observation may be informative: type A side effects tend to occur early in the study (although this period may also be affected by carryover effects from previous

Table 14.3. All events reported on green forms for meloxicam

EVENT	Total	Mth 1	Mth 2	Mth 3	Mth 4	Mth 5	Mth 6	Mths 1-6	Not known
Denominator total	130 615	19 083	19 075	19 068	19 063	19 054	19 046		
Denominator male	41 711	6 172	6 167	6 166	6 164	6 160	6 158		
Denominator female	86 690	12 586	12 583	12 577	12 574	12 568	12 562		
Skin									
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	1
Eczema	71	10	9	8	9	5	7	48	1
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	—

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.

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 14.5 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 comparative study. The ranked reasons for withdrawal can be compared with the ranked incidence density estimates and this comparison can be signal generating. There is usually a good correlation, in terms of the most frequently reported events, and an example of this is given in Table 14.6; other examples have been published elsewhere.⁹⁻¹¹

Comparison of Event Rates and Adjusted Rates

Rate comparisons can be helpful in exploring apparent associations. An example¹² occurred when looking at the long term safety of oral

compared with inhalational β -agonists. The original signal detected in this study was that the value of ID_1 with bambuterol for the event "cardiac failure" was 4.9 per thousand patient months of exposure, compared with 0.6 for salmeterol and 0.4 for nedocromil. Why was this value for bambuterol over eight times higher than that for salmeterol and nedocromil?

Further analysis showed that the relative risk of nonfatal cardiac failure adjusted for age, sex, indication, and season of starting treatment with bambuterol was 3.31 (95% confidence limits 1.91 to 5.71) when compared with nedocromil. When salmeterol was compared with nedocromil, the adjusted relative risk of nonfatal cardiac failure was 0.90 (limits 0.50 to 1.60). Such highly suggestive differences clearly need further exploration.

Other signals and comparisons continue to be explored. These include hallucinations associated with tramadol,¹³ gastrointestinal intolerance with acarbose,¹⁴ galactorrhoea with moclobemide,¹⁵ esophageal reactions with alendronate,¹⁶ a comparison of five antibiotics,¹⁷ and a comparison of antidepressant drugs.¹⁸

Long Latency Adverse Reactions

Special interest attaches to 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 finasteride¹⁹, 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 gynecomastia were only rarely received before the fifth month of therapy. A further important example^{20,21} has occurred in 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 followup 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 lesion has risen to some 2% and that many of the relevant patients have objective evidence of visual field defects.

Outcomes of Pregnancy

Special interest attaches to determining the proportion and nature of congenital anomalies in babies born to women exposed to newly marketed

Table 14.5. Incidence densities (IDs) ranked for meloxicam in order of ID₁ per 1000 patient months

Event	N ₁	N ₂	ID ₁	ID ₂	ID ₁ - ID ₂	99% min	CI max	N _A	ID _A
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
3 Respiratory tract infection	214	387	13.9	8.7	5.2	2.5	7.9	675	9
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	357	4.8
6 Diarrhoea	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	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	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
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
16 Constipation	56	57	3.6	1.3	2.4	1	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	2.1	0.9	3.3	107	1.4
20 Rash	46	90	3	2	1	-0.3	2.2	151	2
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	-0.1	-1.1	1	142	1.9
28 Dyspnoea	29	38	1.9	0.9	1	0.1	2	73	1
29 Fall	26	77	1.7	1.7	0	-1	1	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	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 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 Ischaemic heart disease	24	43	1.6	1	0.6	-0.3	1.5	78	1
36 Cardiac failure	24	29	1.6	0.7	0.9	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	0.4	-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	1.6	46	0.6
41 Orthopaedic surgery	19	95	1.2	2.1	-0.9	-1.8	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	0.9	0.2	-0.6	0.9	61	0.8
44 Anxiety	16	34	1	0.8	0.3	-0.5	1	58	0.8

drugs during the first trimester. PEM studies have shown²² that from 831 such pregnancies, 557 infants were born, of whom 14 (2.5%) had congenital anomalies. This work has not shown unexpected findings, but continues in order to exclude, to the greatest extent possible, unexpected teratogenic effects.

Table 14.6. Reasons for stopping meloxicam (higher level terms)

Reason	Number
Not effective	2989
Condition improved	1834
Dyspepsia	539
Nausea, vomiting	209
Pain abdomen	171
Noncompliance	117
Gastro-intestinal unspecified	104
Diarrhea	103
Orthopaedic 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
Dyspnoea	16
Haemorrhage gastro-intestinal unspecified	16
Surgery unspecified	16
Tinnitus	16
Distension abdominal	12
Drowsiness, sedation	12
Haemorrhage gastro-intestinal upper	12
Pain chest, tight chest	12
Anemia	11
Pain joint	11
Hemorrhage rectal	10

STUDIES OF BACKGROUND EFFECTS AND DISEASES

Background Effects

A database the size of that accumulated in the study of 65 drugs with an average cohort size of over 10 000 patients obviously provides an opportunity to study issues other than the tolerance and toxicity of each specific drug. An issue which has attracted our attention is that the values of ID₁ and ID₂ for upper respiratory tract infection vary very little throughout most PEM studies.²³ This effect is apparent in the data of 40 PEM studies²⁴, and suggests that cohorts have colds in the nose month in and month out, and this provides a rough measure of the background incidence of a disease which, with many drugs, is independent of both the illness being treated and the drug being used for treatment. Seasonal effects are, of course, marked with some drugs, such as antihistamines and antibiotics, but with many drug classes the ID values for upper respiratory infection will show what effects are more frequently seen, or less frequently seen, than this simple clinical yardstick of the rate of URTIs.

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 spontaneously reported an event as an ADR (in a patient being monitored by PEM). It has been shown²⁵ that of 3045 ADRs reported on the PEM green form questionnaires, only 275 (9.0%) had also been reported on yellow cards to the Committee on Safety of Medicines. It is of great interest that 26 (32.1%) of 81 serious unlabelled (i.e., not listed in the Data Sheet/Summary of Product Characteristics for the drug) reactions had been reported on yellow cards. This suggests that doctors use the spontaneous adverse reaction reporting scheme more energetically when reporting those unexpected reactions that worry them most. Our studies in this area have also shown²⁶ 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. This sex difference occurs in all age groups over 19 years of age.

Study of Diseases, Bias, etc.

All discussions of PEM studies need to emphasize that the events reported may be due to chance, the disease being treated, concurrent diseases, concomitantly administered drugs, over-the-counter medication, carryover effects of previous medication, channelling, switching and a whole panoply of confounding and biasing effects—all of which need to be considered before it can be assumed that the drug being monitored might be causing the event being reported.

The database does, however, allow study of diseases as well as drugs, and current studies are concerned with the serotonin syndrome and Churg–Strauss syndrome. In a study of 622 294 patients observed during 58 completed PEM studies, there were four cases of Churg–Strauss syndrome identified after followup. This gives an overall period prevalence estimate of 6.8 (95% confidence limits: –1.8 to 17.3) per million patient years of observation.

An example of the epidemiological aspects that have been studied is a paper on cough with ACE inhibitors—a study of reporting bias in observational cohort studies.²⁷

THE FUTURE

Current efforts, other than continuing an ongoing program of 12 or more PEM studies being conducted simultaneously, concern the validation of PEM and its extension into hospital practice.

Much time and staff resource is also being expended in strengthening the Drug Safety Research Unit's capability to undertake the formal methods of epidemiology, such as case–control studies, comparative cohort studies, and randomized controlled clinical trials. There is only limited point in being able to generate signals of possible adverse drug effects, unless these can be confirmed or refuted by appropriate studies.

Plans for the future largely concern our development of methods, based upon PEM, to study the comparative efficacy and safety of new drugs and their cost-effectiveness.

ACKNOWLEDGEMENTS

I wish to thank the vast number of general practitioners who have taken part in these studies. I am also most grateful for the cooperation of the Prescription Pricing Authority and the NHS authorities. My colleagues and collaborators in these studies have been Dr Lynda Wilton, Dr Nick Dunn, Dr Fiona MacKay, Dr Richard Martin, Mrs Gill Pearce, and Mr Shayne Freemantle. I also wish to thank Mrs Georgina Spragg for administrative assistance.

REFERENCES

1. McBride WG. Cited in Mann RD. *Modern Drug Use—An Enquiry on Historical Principles*. Lancaster: MTP Press–Kluwer, 1984; 606.
2. Committee on Safety of Drugs. Cited in Mann RD. *Modern Drug Use—An Enquiry on Historical Principles*. Lancaster: MTP Press–Kluwer, 1984; 619.
3. Rawlins MD, Jefferys DB. Study of United Kingdom product licence applications containing new active substances, 1987–9. *Br Med J* 1991; **302**: 223–5.
4. Rawlins MD, Mann RD. Monitoring adverse events and reactions. In: *A Textbook of Pharmaceutical Medicine, Current Practice*. Mann RD, Rawlins MD, RM Auty (eds), Carnforth: Parthenon, 1993; 319.
5. Dunlop D. Cited in Mann RD. The ethos of drug safety monitoring. *Br Med J*. 1998; **316**: 837.
6. Committee on Safety of Medicines. Cited in *Adverse drug reactions*. (Ed: RD Mann). Carnforth: Parthenon, 1987, pp. 62–3.
7. Bem JL, Breckenridge AM, Mann RD and Rawlins MD. Review of yellow cards (1986): report to the Committee on the Safety of Medicines. *Br J Clin Pharmacol* 1988; **26**: 679–89.
8. Freemantle SN, Pearce GL, Wilton LV, Mackay FJ and Mann RD. The incidence of the most commonly reported events with 40 newly marketed drugs—a study by Prescription-Event Monitoring. *Pharmacoepidemiol Drug Safety* 1997; **6** (suppl. 1): 1–8.
9. British Medical Association. *Reporting Adverse Drug Reactions*, a BMA Policy Document. London: British Medical Association, 1996.
10. Mann RD, Wilton LV, Pearce GL, Mackay FJ and Dunn NR. Prescription-Event Monitoring (PEM) in 1996—a method of non-interventional observational cohort pharmacovigilance. *Pharmacoepidemiol Drug Safety* 1997; **6** (suppl. 3): S5–11.

11. Mann RD. Prescription-Event Monitoring—recent progress and future horizons. *Br J Clin Pharmacol* 1998; **46**: 195–201.
12. Martin RM, Dunn NR, Freemantle SN and Mann RD. Risk of non-fatal cardiac failure and ischaemic heart disease with long acting beta₂ agonists. *Thorax* 1998; **53**: 558–62.
13. Dunn N, Wilton LV, Pearce G and Mann RD. The incidence of hallucinations associated with tramadol. *Pharmacoepidemiol Drug Safety* 1997; **6** (suppl. 2): S19.
14. Mackay F, Wilton LV, Pearce G, Freemantle S and Mann RD. Acarbose and gastrointestinal intolerance. *Pharmacoepidemiol Drug Safety* 1997; **6** (suppl. 2): S19.
15. Dunn NR, Freemantle SN, Pearce GL and Mann RD. Galactorrhoea with moclobemide. *Lancet* 1998; **351**: 802.
16. Mackay FJ, Wilton LV, Pearce GL, Freemantle SN and Mann RD. United Kingdom experience with alendronate and oesophageal reactions. *Br J Gen Pract* 1998; **48**: 1161–2.
17. Wilton LV, Pearce GL and Mann RD. A comparison of ciprofloxacin, norfloxacin, ofloxacin, azithromycin and cefixime examined by observational cohort studies. *Br J Clin Pharmacol* 1996; **41**: 277–84.
18. Mackay FJ, Dunn NR, Wilton LV, Pearce GL, Freemantle SN and Mann RD. A comparison of fluvoxamine, fluoxetine, sertraline and paroxetine examined by observational cohort studies. *Pharmacoepidemiol Drug Safety* 1997; **6**: 235–46.
19. Wilton L, Pearce G, Edet E, Freemantle S, Stephens MDB and Mann RD. The safety of finasteride used in benign prostatic hypertrophy: a non-interventional observational cohort study in 14 772 patients. *Br J Urol* 1996; **78**: 379–84.
20. Stephens MDB, Wilton LV, Pearce G and Mann RD. Visual field defects in patients taking vigabatrin. *Pharmacoepidemiol Drug Safety* 1997; **6** (suppl. 2): S18.
21. Wilton LV, Stephens MDB and Mann RD. Visual field defects in patients on long term vigabatrin therapy. *Pharmacoepidemiol Drug Safety* 1999; **8**(suppl. 2): 5108.
22. Wilton LV, Pearce GL, Martin RM, Mackay FJ and Mann RD. The outcomes of pregnancy in women exposed to newly marketed drugs in general practice in England. *Br J Obstet Gynaecol* 1998; **105**: 882–9.
23. Wilton LV, Freemantle SN, Martin R and Mann RD. Is the incidence of upper respiratory tract infection independent of drug treatment in large cohort studies of longer term use drugs? *Pharmacoepidemiol Drug Safety* 1998; **7**: S4–10.
24. Freemantle SN, Pearce GL, Wilton LV, Mackay FJ and Mann RD. The incidence of the most commonly reported events with 40 newly marketed drugs—a study by Prescription-Event Monitoring. *Pharmacoepidemiol Drug Safety* 1997; **6** (suppl. 1): 9–62.
25. Martin RM, Kapoor KV, Wilton LV and Mann RD. Underreporting of suspected adverse drug reactions to newly marketed (“black triangle”) drugs in general practice: observational study. *Br Med J* 1998; **317**: 119–20.
26. Martin RM, Biswas PN, Freemantle SN, Pearce GL and Mann RD. Age and sex distribution of suspected adverse drug reactions to newly marketed drugs in general practice in England: analysis of 48 cohort studies. *Br J Clin Pharmacol* 1998; **46**: 505–11.
27. Wilton LV, Pearce G and Mann RD. Cough with ACE inhibitors: an example of reporting bias in observational cohort studies. *Pharmacoepidemiol Drug Safety* 1996; **5** (suppl. 1): S13.
28. Finney DJ. The design and logic of a monitor of drug use. *J Chron Dis* 1965; **18**: 77–98.

15

Group Health Cooperative of Puget Sound

KATHLEEN W. SAUNDERS

The Center for Health Studies, Group Health Cooperative of Puget Sound, Seattle, WA, USA

ROBERT L. DAVIS

The Center for Health Studies, Group Health Cooperative of Puget Sound, Department of Pediatrics, Department of Epidemiology, University of Washington, Seattle, WA, USA

ANDY STERGACHIS

Department of Pharmacy, Department of Epidemiology, University of Washington, Drugstore.com, Bellevue, WA, USA

INTRODUCTION

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 healthcare within these integrated systems through a uniform benefit package.¹ Care is usually provided within a defined geographic area that allows for large comparable groups of subjects for public health research.¹

The unique features of traditional group and staff model HMOs are being diluted somewhat by the pressure of market forces. “Managed care” organizations (MCOs) have been identified as the solution to increasing healthcare costs¹ and enrollment in MCOs is now more than 70% of the US market.² While the term “managed care” was once synonymous with traditional HMOs, it now encompasses new organizational forms such

as network model HMOs and individual practice associations (IPAs).¹ Most of these new models feature insurance organization contracts with physicians’ groups to provide care for enrollees, as opposed to fully integrated delivery systems.¹ In order to compete with these new healthcare models, staff model HMOs are using more mixed models of care, creating their own networks and IPAs.¹ Pressure is also mounting to offer a wider selection of benefit options, such as point-of-service plans, which allow enrollees to seek care wherever they want.¹ “One-size-fits-all” benefit packages that used to be the rule in HMOs have been transformed into individualized plans that may number in the hundreds. Thus, while traditional HMOs continue to play an important role in public health research, including post-marketing drug surveillance, it is important to keep in mind the market trends that threaten some of the unique features of these organizations.

Some of these trends are evident at Group Health Cooperative of Puget Sound. While this chapter emphasizes the advantages of conducting postmarketing drug surveillance in a predominantly closed staff-model HMO, it also acknowledges the implications of moving toward a more mixed model of care.

There has been longstanding interest in the use of data from HMOs to study the effects of marketed drugs. Since the time of the report of the Joint Commission on Prescription Drug Use, recommendations have been made on the use of HMO records for postmarketing drug surveillance.³ There are several advantages to conducting research in an HMO setting.⁴ 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. 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.

Data from the Group Health Cooperative of Puget Sound (GHC) have been used extensively to evaluate drug usage and the adverse and beneficial effects of marketed drugs and medical procedures. In 1983, GHC made explicit its commitment to research and evaluation by establishing the Center for Health Studies. The mission of the Center for Health Studies is to develop GHC as a setting for population-based and intervention research, through its own program of research and through collaborative ties with other scientists, including those affiliated with the University of Washington, Kaiser Permanente Northwest, and the Fred Hutchinson Cancer Research Center.

GHC data used in postmarketing drug surveillance research are generally derived from automated files containing information recorded during the routine delivery of health services. These data have been used to identify prospective and historical cohorts of users of drugs or other exposures of interest, appropriate comparison groups, clinical events, health services utilization, and confounding variables. This chapter reviews the characteristics of the GHC setting and

databases, the advantages and disadvantages of GHC data resources for postmarketing drug surveillance, selected methodologic issues pertaining to postmarketing drug surveillance that arise in an HMO setting such as GHC, and expected future directions for research with the GHC databases.

DESCRIPTION

GROUP HEALTH COOPERATIVE OF PUGET SOUND

GHC, a nonprofit consumer-directed HMO established in 1947, provides health care on a prepaid basis to approximately 415 000 persons in western Washington State. Most of these are "staff model" enrollees who receive all care at Group Health facilities, with the exception of specific services not provided by Group Health providers (e.g., temporomandibular care). In the latter case, the Cooperative contracts with selected community providers to whom enrollees are referred. Approximately 15% 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. About 8% of enrollees have chosen point-of-service plans, which allow them to seek care wherever they want.

In addition to these 415 000 GHC enrollees, approximately 47 000 Western Washington residents belong to Options Health Care, Inc., a wholly owned subsidiary of GHC established in 1990. These "point-of-service" enrollees can receive care from Group Health providers at greater benefit coverage, or from community providers with less coverage. In 1996, only 17% of Options enrollees' healthcare costs were incurred through these "outside" community providers.

Approximately 13% of Western Washington residents were enrolled in GHC in 1998. The majority of GHC enrollees, approximately 80%, receive health benefits through their place of employment (i.e., group enrollees). At present, nearly 2000 Puget Sound firms and unions offer GHC coverage to their employees. In addition, as

of January 1999 GHC had arrangements for providing services to approximately 54 000 Medicare, 25 000 Medicaid, and 26 000 Washington Basic Health Plan recipients. The Basic Health Plan is a state subsidized program that provides medical insurance to low income, uninsured residents who earn too much to qualify for Medicaid. In the last 6 years, the number of GHC Medicaid enrollees has increased more than tenfold. This increase is due to the advent of Healthy Options, the State of Washington's Medicaid managed care program. GHC is one of the healthcare contractors involved in this program.

For the most part, GHC offers comprehensive health care coverage for outpatient care, inpatient services, emergency care, mental health services, and prescribed drugs. (A major exception is the lack of drug coverage for Medicare enrollees new to GHC since 1994.) As mentioned above, this healthcare is provided primarily through GHC's own facilities, including one hospital, one skilled nursing facility, 26 primary care clinics, and four specialty centers. GHC operated two hospitals until the Cooperative formed a strategic alliance with Virginia Mason Medical Center in 1994, which resulted in the closure of one hospital; the two organizations contract with each other for hospital services. In 1999, nearly all benefit plans required small copayments for services, such as prescriptions, nonpreventive care outpatient visits, and emergency treatment. Coverage policies for outpatient drugs are controlled by GHC's drug formulary.

GHC contracts with Group Health Permanente, a partnership of physicians responsible for providing medical services at the Cooperative's facilities. In 1998, this Permanente Medical Group included 550 physicians, approximately 90% of whom were board certified. The large majority of GHC's primary care physicians were trained in family practice residencies. GHC has a formal affiliation agreement with the University of Washington School of Medicine, the first in the nation between an HMO and a major medical school. The distribution of patients to primary care providers' practices is based on a panel system in which each primary care provider

(family practice physician, pediatrician, or internist) has responsibility for managing and coordinating the care of a panel or caseload of patients. Upon enrollment in GHC, patients are offered a choice of primary care physicians and may change primary care physicians at any time during their tenure in the plan. More than 95% of GHC enrollees are identified with a specific panel. In 1998, GHC enrollees had nearly 4 million outpatient drug prescriptions and almost 2.5 million outpatient visits to GHC physicians or allied professionals.

As shown in Table 15.1, compared to other Seattle-Tacoma area residents, GHC enrollees have slightly higher educational attainment but are quite 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 population (fewer blacks, higher educational level, less representation within the lowest extreme of income distribution among GHC enrollees) primarily reflect differences between the demographic composition of Seattle-Tacoma and the US population as a whole.

DATABASES AT GHC

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 dis-enrolls and rejoins GHC at a later date.

Table 15.2 lists the research data sets that have been developed from GHC's databases. Data are available in SAS format on several different platforms such as an MVS mainframe, a UNIX system, and the Center for Health Studies' Data Warehouse on an NT server. For relational database purposes, much of these data are available in a Sybase Data Warehouse on a UNIX system. Files are updated daily, monthly, quarterly, or semi-annually, using data from clinical and administrative computer systems. Every file

Table 15.1. Demographic Characteristics of Adult Group Health Cooperative Enrollees

Characteristic	Percentage of total adult population		
	GHC enrollees ^a	Seattle–Tacoma area ^b	United States ^c
Age (years)			
18–44	57	62	58
45–64	27	24	25
65+	16	14	17
Sex			
Male	45	49	48
Female	55	51	52
Race or ethnicity			
White	90	88	82
Black	3	4	11
Asian or Pacific Islander	4	6	3
American Indian, Eskimo or Aleut	1	1	1
Other	2	1	3
Years of education			
<12	9	14	25
12	24	27	30
13–15	33	34	27
>15	34	25	19
Annual income ^d			
<\$15 000	17	17	24
\$15–30 000	28	24	26
\$30–45 000	26	23	21
\$45–60 000	16	15	13
>60 000	13	20	17

^aAge and sex based on 1990 Group Health enrollment data. Race based on two random samples of Group Health enrollees surveyed in 1990: 1308 enrollees age 18–64²⁵ and 2513 enrollees age 65 and up.⁵⁸ Years of education and household income based on a random sample of 1133 enrollees surveyed in 1984.²³

^bBased on 1990 Seattle-Tacoma CMSA.

^cBased on 1990 census.

^dIncome figures are corrected for inflation between 1984 and 1989 and stated in 1989 dollars.

contains the unique patient identifier common to all of the datasets. Physician identifiers are also unique across all files. A brief description of each of GHC's data files is found in Table 15.2.

Enrollment

There are two different types of enrollment files—current and historical. The current enrollment file

Table 15.2. Automated databases organized for analytic studies at GHC

Database	Years Available
Enrollment	All
Pharmacy	March 1977–present
Hospitalization	January 1972–present
Laboratory	January 1986–present
Radiology	January 1986–present
Outpatient visits without diagnosis	July 1984–December 1991
Outpatient visits with diagnosis	January 1992–present
Cancer surveillance system	January 1974–present
Cause of death	January 1977–present
Community health services	June 1989–present
Cost Information System	June 1989–present
Claims from non-GHC providers	June 1989–present
Breast cancer screening program	January 1985–present
Immunization	February 1991–present
Diabetes registry	1996–present
Heart care registry	1997–present
Selected population-based surveys	
Health risk survey	1984
Senior survey	1985
Chronic pain	1986
Tobacco use	1991–1992

consists of a record for every person currently enrolled in GHC, some 460 000 enrollees. It contains 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. This file is often used to select probability samples of current GHC enrollees. The second type of enrollment file is historical. One dataset contains a record for each of the roughly 2 000 000 persons who have ever been enrolled at GHC. This dataset includes a person's estimated original enrollment date (not available for all persons). A second historical dataset contains all enrollment periods (up to five) for the nearly 1.5 million persons who have been enrolled at GHC at any time since 1980.

Pharmacy

The pharmacy file includes data on each prescription dispensed at GHC-owned outpatient pharmacies since March 1977. Table 15.3 lists the five most frequently dispensed drugs in each of nine major drug categories in 1998. This table includes over-the-counter as well as prescription medicines. A computerized record is created for every medication at the time the prescription is filled. As shown in Table 15.4, each record contains selected information about the patient, the prescription, and the prescriber. A new variable, "dayssply", was added to the pharmacy database in 1996. This field, motivated by the Cooperative's need to charge the appropriate one-, two-, or three-month copay amount, indicates the number of days the medication should last.

Hospital

The hospitalization databases contain records of every discharge, including newborn and stillborn infants, from GHC-owned hospitals since January 1972. The Commission on Professional Hospital Activities—Professional Activity Study (CPHAS) provided hospital discharge information from January 1972 through December 1984. Since January 1985, GHC's hospital information system (HIS) has included information for discharges from GHC-owned hospitals and a GHC-operated wing of a community hospital. The information in the hospitalization database includes patient characteristics, diagnoses, procedures, diagnostic related group (DRG), and discharge disposition. While data on inpatient prescription drug usage became available in mid-1989, this aspect of the HIS system has not been used extensively for analytic purposes. Virtually all hospitalizations not captured in HIS are included in GHC's outside claims files. To facilitate evaluation of hospital costs and utilization, all records of discharges from GHC and non-GHC hospitals since June 1989 have been combined into an annual utilization dataset.

Laboratory

Automated laboratory data are available from January 1986. The on-line 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 the name of the test ordered, the date ordered, the specimen source, the results, and the date of the results. This database has been used in studies of acyclovir safety⁵ and studies of the epidemiology of *Chlamydia trachomatis* infection.⁶ This database has also been used to evaluate the implementation of new Centers for Disease Control (CDC) guidelines on preventing Group B Streptococcal (GBS) disease. GHC laboratory data provided information on GBS screening during routine prenatal care visits and allowed GHC and the CDC to evaluate the rapidity of uptake of the new guidelines. The laboratory data also allowed investigators to evaluate how completely the new guidelines were being implemented.⁷

Radiology

This system, established in 1986, contains records of all radiographic studies performed at GHC facilities, including CT and MRI scans. Radiology records were maintained in a separate database through June 1995; since then these records have been combined with those in the outpatient visits file.

Outpatient Visits

The GHC outpatient registration system was initiated in mid-1984 and includes selected information about each outpatient visit. The database includes a record of the date of visit, the provider seen, the provider's specialty, and the location of care. Beginning in 1991, and fully operational in 1992, diagnosis and procedure data have been incorporated in the registration database. This information was not available in automated form before that time. This database has been used to examine ambulatory utilization patterns related to quitting smoking⁸, ambulatory utilization patterns of children of smokers,⁹ and back pain related visit utilization following randomization to one of three treatments for low back pain.¹⁰ Outpatient visit data have been particularly useful in studies of

Table 15.3. Frequently dispensed drugs at GHC outpatient pharmacies in each of nine major drug categories: January–December 1998^a

Major category	Drug	No. of people	No. of prescriptions
Analgesics	Acetaminophen	44 565	107 888
	oxycodone/hydrocodone		
	Acetaminophen	25 974	44 771
	Acetaminophen–codeine	24 920	48 314
	Aspirin	21 151	51 608
Antihypertensives	Propoxyphene combination	6 082	15 983
	Atenolol	19 625	77 562
	Lisinopril	19 472	89 151
	Propranolol	2 677	9 129
	Prazosin	2 638	11 028
Anti-infectives	Clonidine	1 554	6 728
	Amoxicillin	59 642	83 375
	Trimethoprim-sulfamethoxazole	39 601	53 809
	Cephalexin	34 184	46 175
	Erythromycin	16 415	21 139
Nonsteroidal anti-inflammatory drugs	Doxycycline	14 456	20 142
	Ibuprofen	43 714	75 467
	Naproxen	25 188	45 471
	Sulindac	5 424	12 072
	Indomethacin	4 264	8 296
Calcium channel blockers	Piroxicam	2 228	5 820
	Felodipine	4 933	26 796
	Verapamil	4 565	18 910
	Diltiazem	3 136	13 922
	Amlodipine	645	3 189
Diuretics	Cardizem	511	2 176
	Hydrochlorothiazide	14 896	49 122
	Furosemide	9 610	40 529
	Hydrochlorothiazide–triamterene	7 498	27 661
	Hydrochlorothiazide–spironolactone	1 492	4 939
Gastrointestinal	Spironolactone	1 106	3 806
	Cimetidine	18 121	47 418
	Ranitidine	12 404	42 855
	Docusate sodium	8 525	16 686
	Lansoprazole	5 784	23 785
Psychotherapeutics	Antacids	5 611	11 401
	Paroxetine	13 886	52 231
	Trazadone	8 745	33 156
	Lorazepam	6 888	23 257
	Fluoxetine	6 635	34 297
Respiratory agents	Sertraline	6 282	27 628
	Nasacort	37 621	79 821
	Albuterol	33 951	82 153
	Fluticasone	3 971	13 505
	Ipratropium	2 435	9 367
	Beclomethasone	2 035	5 229

^aDoes not include prescriptions from non-GHC pharmacies.

Table 15.4. Selected variables available from automated drug files on all prescriptions dispensed from GHC outpatient pharmacies, 1977 to present

Patient Data
Consumer No.
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
Pharmacy location

antibiotic usage among GHC children, because the majority of pediatric antibiotic use occurs in the context of outpatient ambulatory visits.¹¹

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 tapes 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 United States.¹² The CSS reporting area consists of the 13 contiguous counties of northwest Washington State. The database, which currently covers the period 1974 through 1998, contains information for each newly diagnosed cancer case, including patient demographics, anatomical site, stage at diagnosis, and vital status at followup, which is ongoing for all surviving cases in the register. This database has been used to evaluate GHC's Breast Cancer Screening Program¹³ and mammography performance.¹⁴

GHC has also developed diabetes and heart disease registries as part of its commitment to "population-based care".^{15,16} This concept refers to *initiating* systematic care for groups of patients with shared healthcare needs (e.g., all diabetics), instead of *reacting* to individual patient's crises (the acute care model). The first step in implementing population-based care is to develop registries to identify the relevant population of patients. In the case of diabetes and heart disease, GHC has moved from paper-and-pencil to automated registries.

Cause of Death

Using data from the State of Washington vital statistics files, the Cancer Surveillance System, and GHC's hospitalization files, a file of deaths among GHC enrollees between the years 1977 and 1997 has been developed. The death file was produced through record linkages between the State of Washington's death certificate database and the GHC membership files.

Community Health Services System

GHC's Community Health Services department operates visiting nurse, hospice, respite, and geriatric nurse practitioner programs, a nursing home rounding system, and numerous other community based service programs. Computerized information from this department, available since mid-1989, includes type of provider, type of procedure, diagnosis, location of service, and price and billing information.

Utilization Management/Cost Management Information System (UM/CMIS)

Due to a growing need for utilization and cost information, the Utilization Management/Cost Management Information System (UM/CMIS) was developed at GHC in 1989. This system uses information from feeder systems (e.g., registration, pharmacy) to assign both direct and indirect costs to individual encounters based on the units of service utilized. Thus, it is possible to use UM/CMIS data 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. These data have been used to study the healthcare costs associated with several conditions or diseases, including asthma,¹⁷ depression,¹⁸ and breast cancer.¹⁹ Analyses related to costs of care for children with attention deficit disorder are underway.

Immunizations Database

The immunizations database was initially developed as part of the Vaccine Safety Datalink (VSD), a CDC collaborative study with GHC, Kaiser Permanente Northwest, Northern California Kaiser, and Southern California Kaiser. 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. The immunization database has been used in studies of vaccine safety²⁰ and in studies of the utility of immunization tracking systems.^{21,22}

Other Databases

Claims databases record information on health care services which GHC purchases from non-GHC providers. These databases contain a record of each bill received by GHC for outside services since 1989, such as inpatient and emergency care and referrals to 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 means of giving a full accounting of an individual's medical care. For example, they capture bills from network providers who have contracted to provide care to GHC enrollees in outlying areas and from community providers who deliver care to enrollees in point-of-service plans.

Important data are also being generated from an ongoing program on breast cancer screening. All GHC women aged 40 years or older are being mailed questionnaires that solicit information on factors associated with the incidence of breast cancer. Approximately 85% of the women complete

the questionnaires. Programmatically, the information from this database is being used to ascertain individual level of risk for breast cancer, with a woman's risk group determining the frequency of screening mammography. Included in the breast cancer screening database are items pertaining to lifetime smoking history, use of estrogens, and reproductive history.

Data obtained from population based GHC surveys have also served as resources for research studies. In many instances, survey data have been linked with automated files to investigate predictors of health care utilization. The Health Risk Survey data set consists of responses from an age and sex stratified random sample of 1133 GHC adult enrollees.²³ The survey form included standardized questions about health status, socio-demographic information, and health-related practices. Another example is a survey of a random sample of 647 GHC enrollees aged 65 years and older. Among the topics included in this Senior Survey were health status, functional status, situational status, health-related practices, and sociodemographics. Other survey data have focused on specific topics such as chronic pain²⁴ and tobacco use.²⁵

A variety of 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 460 000 persons, and this allows the study of relatively uncommon adverse events. The GHC population is well defined in terms of age, sex, geographic location, and duration of GHC membership. Turnover in membership at GHC is estimated to be approximately 15% per year. This relative stability facilitates long-term

follow-up—Center for Health Studies' longitudinal studies routinely obtain followup data for 85% of the original cohort after a one-year period. Typically, such studies are successful in following persons by telephone whether or not they remain GHC enrollees.

A single patient identification number permits following the records of individuals across time. The unique patient consumer number is assigned to each enrollee upon initial enrollment at GHC and remains with the individual even if disenrollment and re-enrollment occurs.

An obvious advantage to the GHC setting for research is the extensive use of computer technology. As previously noted, numerous automated databases are available at GHC and can be utilized for postmarketing drug and vaccine safety surveillance. 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. Thus, the high cost and potential for recall bias that would otherwise be associated with primary data collection are minimized or avoided. However, GHC has mechanisms for contacting patients directly for primary data collection, as appropriate. Examples include interview based case-control studies, health surveillance activities, and clinical research studies. In addition, manual medical records and registries are also available for use.

Another advantage to GHC as a research setting is the longevity of the plan. In existence for more than 50 years, Group Health has a subset of enrollees whose tenure in the Cooperative spans decades. This longevity facilitates studies that require very long term followup. For example, one study examined a hypothesized relationship between inflammatory bowel disease (IBD) and measles vaccination. The onset of IBD often occurs in late adolescence or early adulthood, yet the immunization of interest (MMR vaccine) is delivered in the second year of life. This long lag between exposure and outcome necessitated the creation of a dataset of members born into GHC as far back as 1958 who remained enrolled until at least 1972. Medical records (both hospitalization and outpatient) were used to confirm the diagnosis of IBD, and, since these records extended back to

birth, were also used to collect information on childhood immunizations.

WEAKNESSES

LOGISTIC AND OPERATIONAL

Despite the large size of the GHC databases, most drugs are used by a relatively small proportion of a population. Thus, the GHC databases still may be too small to detect associations between drugs and rare outcomes. The detection of rare adverse events requires combining data from multiple health care delivery systems.

The data currently available in automated files do not include some potentially important confounding variables. The lack of information on confounding factors such as smoking and alcohol consumption may lead to invalid conclusions regarding drug associated effects. Confounding by indication occurs when the underlying diagnosis or other clinical features that trigger use of a certain drug also predict patient outcome in their own right. For example, in a study of β -blockers and the risk of coronary heart disease (CHD) among persons with hypertension, confounding by indication was a concern because β -blockers are also used to treat angina pectoris and are avoided in cases of congestive heart failure, each of which may be an early manifestation of CHD.²⁶ Analytic studies must consider alternative methods of obtaining the necessary information on confounding factors, such as medical record abstraction and patient interview. Also, automated information on inpatient drugs and outpatient diagnoses were not available until relatively recently, an important disadvantage for some retrospective studies.

Use of automated data to determine outcome is not always reliable without chart review. In the Vaccine Safety Datalink study of seizures following vaccination, it would have been optimal to rely on computerized seizure related diagnosis codes. However, the investigators found that these codes often identified visits of children with seizure conditions who were being seen for followup or well child care. They concluded that for chronic conditions such as seizure disorders or asthma,

medical record review is often necessary to distinguish acute events from followup or routine visits.²⁷ In addition, "rule out" diagnoses can be misleading. In reviewing the medical records of a subset of people receiving a computerized diagnosis of rheumatoid arthritis (RA), one GHC researcher found that it was not uncommon for the diagnosis of RA to have been subsequently ruled out by further medical tests.

The present competitive environment in the health care industry has resulted in a growing number of mixed model benefit plans (e.g., non-GHC provider networks and point-of-service plans) being offered by GHC, which may jeopardize the completeness of databases of health services utilization. However, if necessary, it is possible to use automated data to exclude "non-staff" model enrollees from studies. 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 (e.g. State employees, BHP enrollees). For example, since 1994, GHC has not provided drug coverage to new Medicare enrollees (see "completeness" section). 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, changes that could have serious ramifications for healthcare utilization. Increased competition may also lead to increased patient turnover in HMOs, resulting in decreases in followup time for cohort studies. The costs to the HMO associated with maintaining relatively complete healthcare data and archiving data for analytic use may also become an issue as the pressure for health care cost containment continues.

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 one brand of a drug on the formulary at a time, thereby preventing investigations of many direct drug-to-drug comparisons of relative toxicity and effectiveness. In particular, drugs which offer little therapeutic advantage or

value over alternative agents may be excluded from the GHC formulary. If nonformulary drugs are commonly purchased outside the GHC pharmacy system, as may have been the case with Sildenafil (Viagra) and Fen/Phen, 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.

The poor and elderly have tended to be underrepresented in HMOs,¹ leading to concerns about representativeness of studies. However, a Medicare managed care plan has been offered since 1985; GHC's enrollee population is currently made up of a greater proportion of Medicare recipients than is the population of the Seattle Metropolitan area.¹ Group Health's involvement in Healthy Options, Washington State's Medicaid managed care program, has resulted in an increase in Medicaid enrollment from 2000 to 25 000 between 1993 and 1998.

PARTICULAR APPLICATIONS

EXAMPLES OF USE OF GHC DATA

The principal use of the pharmacy database for epidemiologic research has been to ascertain drug exposures, often for evaluating the effectiveness or toxicity of specific medications. Examples of questions addressed by case-control studies include: (i) do β -blockers reduce the incidence of coronary heart disease in patients with high blood pressure;²⁸ (ii) are thiazide diuretics associated with hip fracture;²⁹ and (iii) is the use of exogenous hormones associated with increased risk for temporomandibular pain?³⁰ The database has also been used in retrospective and prospective cohort studies of specific medications. Examples include studies of the medical outcomes and costs associated with pentoxifylline treatment of patients with peripheral arterial disease³¹ and the perinatal effects of acyclovir.⁵ Other applications include studying patterns of drug utilization.

Such studies have examined the effect of various levels of prescription drug copayments on overall drug utilization³² and the effect of gaps in mood stabilizer treatment among patients treated for bipolar disorder.³³ One study examined physician variation in prescribing rates for opioid medications. Opioid use was determined from self-report after validation against the pharmacy database.³⁴

The pharmacy database is increasingly used as a sampling frame. In some instances, medication use is the primary sampling criterion, as in the case of studies restricted to hypertensive patients treated with medication.³⁵ In other cases, medication use has been used as a proxy measure of disease status.³⁶ For example, one of the methods used to identify cases for a case-control study of the risk factors for pelvic inflammatory disease was to identify women on certain antibiotic regimens (e.g., doxycycline or 2 g or more of tetracycline per day). The medical record was then reviewed to determine if the indication was for pelvic inflammatory disease. Other uses of the database include controlling for potential confounding represented by the use of certain kinds of drugs^{29,37} and drug utilization review activities.³⁸⁻⁴⁰

Below are some detailed examples of how the GHC pharmacy database has been used for epidemiologic and health services research.

Primary care patients treated for depression often discontinue use of antidepressant medications before therapeutic benefit is achieved or take the medications at suboptimal dosages.⁴¹ These treatment deficiencies can be attributed to a combination of patient and provider characteristics and to drug side-effect profiles. GHC and University of Washington researchers conducted a randomized trial testing a patient and provider educational intervention designed to improve adherence to antidepressant therapy.⁴² One outcome measure used in this trial was an automated pharmacy based measure of antidepressant treatment adequacy.⁴³ This measure was based on a computerized algorithm which used the timing and quantity of refills to assess whether a person took an adequate dose of a specific antidepressant medication for a minimum of 30 and 90 days over a 6 month period. The pharmacy-

based measure was also used as an outcome in a randomized trial examining the "real-world" efficacy of three kinds of antidepressant (fluoxetine, imipramine, and desipramine).⁴⁴ A separate analysis found that this automated measure of overall treatment adequacy agreed very well with a self-report measure ($\kappa = 0.8$).⁴⁵

National advisory groups issued conflicting advice as to the timing of the second dose of the measles-mumps-rubella vaccine (MMR2). One group recommended MMR2 at age 4-6 years; a second group recommended that the vaccine occur at age 10-12. As there was little information on the rate of adverse reactions following vaccination in these two age groups, GHC and Northern California Kaiser (NCK) investigators took advantage of differing immunization policies at their respective HMOs to compare the frequency of clinical events following, and possibly related to, MMR2.⁴⁶ The investigators studied over 8000 children receiving MMR2 at age 4-6 years at NCK and over 18 000 children receiving MMR2 at age 10-12 years at GHC. The presence and timing of vaccinations were determined from automated data on vaccines. Automated diagnosis data were used to identify clinical events plausibly associated with the receipt of MMR2 in the two cohorts and then samples of these clinical events were confirmed by chart review.

To account for age-related differences in health-care use, clinical events in a 30 day time period following immunization were compared to those in a 30 day period prior to vaccination. Children 10-12 years of age were 50% more likely to have an adverse event following MMR2 than in the period prior to immunization, while children 4-6 years of age were less likely to have a visit for an event following immunization compared to the period prior to immunization. These results suggested that the risk for adverse clinical events following MMR2 immunizations was greater in the 10-12 year age group.

A population based case-control study assessed the association between first myocardial infarction (MI) and the use of antihypertensive agents.⁴⁷ Investigators hypothesized that hypertensive patients treated with calcium channel blockers, as

opposed to diuretics or β -blockers, may be at increased risk of a first MI. The study covered a time period (1986–1993) when guidelines recommended calcium channel blockers as one of the initial therapies for hypertension.

Cases were defined as GHC enrollees with pharmacologically treated hypertension who suffered a first time fatal or nonfatal MI between 1986 and 1993. Cases were identified from one of three sources: (i) computerized discharge records from two GHC hospitals; (ii) bills for out-of-plan services provided; and (iii) a computerized match between GHC enrollment files and Washington State death registry files. Frequency age–sex matched controls had pharmacologically treated hypertension but did not have an MI.

Analyses were restricted to subjects who were current users of antihypertensive drugs on their index date: the date of MI for cases or, for controls, a computer generated random date within the calendar year for which they had been sampled. To determine current use, the investigators searched the computerized pharmacy database for filled prescriptions immediately preceding the index date. Subjects who received enough pills to last until the index date were considered current users. (Investigators assumed that subjects were at least 80% compliant with dosage instructions.) In order to avoid the potential confounding effects of recently starting drug therapies, investigators also required that the subject took the medication for at least 30 days prior to the index date.

The study found that the use of short-acting calcium channel blockers, especially in high doses, was associated with an increased risk of MI. In part as a result of this study⁴⁷ and in part as a result of a meta-analysis,⁴⁸ the FDA issued a warning against the use of short-acting nifedipine (a calcium channel blocker) in the treatment of hypertension. The Heart Care Team at GHC intervened with physicians to transfer patients off this medication. Within 6 months following the intervention, 80% of patients taking short-acting nifedipine had discontinued use.⁴⁹ Most switched to other drugs, and a small proportion stopped the use of antihypertensives altogether.

METHODOLOGIC ISSUES PERTAINING TO USE OF THE GHC PHARMACY DATABASE

The following sections describe methodologic problems that have been addressed by researchers using the GHC pharmacy database.

Completeness of the Database

An important issue when using GHC's outpatient pharmacy database for postmarketing drug surveillance research is the completeness of the database—that is, what proportion of prescriptions written to GHC enrollees are filled at GHC pharmacies. This issue has become more salient over time because in 1994 Group Health no longer offered prescription medication coverage to Medicare enrollees new to the Cooperative. In 1997, it was estimated that 50% of Medicare enrollees did not have pharmacy coverage. Another factor that could potentially affect the completeness of the database over time is copayments, because reimbursement policies may influence patients' decisions on where they fill prescriptions or obtain health care services. Copayments were introduced for some plans at GHC in 1985, and by 1993 nearly all plans required modest copayments for visits and drugs. Another cause for concern is the increase in the % of enrollees who do not have traditional “staff model” plans. These enrollees receive their prescriptions at community pharmacies. Assuming the enrollee has drug coverage, the community pharmacy then has to bill Group Health in order for the Cooperative to be informed of the fill.

Scientists in the Cardiovascular Health Research Unit regularly ask subjects about the percentage of GHC prescriptions that they purchase at GHC pharmacies. Ninety-seven% of about 2400 control subjects interviewed before 1994 reported that they bought all or almost all (90–100%) of their prescription medications at GHC pharmacies (personal communication with Bruce Psaty). This percentage remained about the same (98%) when asked of 600 control subjects after January 1994 (most were interviewed in 1994 and 1995). Thus, even in a time period (post-1993) when almost all enrollees had requirements for

pharmacy copayments, the pharmacy database appeared to be very complete. Furthermore, Psaty and colleagues were able to examine completeness rates among senior enrollees (65 years of age and older) before and after January 1 1994, the date GHC implemented its policy to no longer cover drugs for new Medicare enrollees. Among subjects 65 years and older, 97.6% filled all or almost all of their prescription medicines at a GHC pharmacy before 1994 compared with 98.3% from January 1 1994 on.

Other survey data also provide information on the completeness of the pharmacy database. According to a survey of pain patients, more than 90% of prescription medications used for pain management, such as opioids, sedatives/muscle relaxants, and anti-inflammatory drugs, were always filled at GHC pharmacies (see Table 15.5). Among 762 study subjects treated with antidepressant medications in 1996–97, only 1.5% reported obtaining antidepressants from a non-GHC pharmacy in the prior 3 months.

Indication for Use

Because there is no variable for disease or symptom indication for prescription in the pharmacy database, the medical record has often been used for this purpose.^{28,31} This brings up two points: (i) to what extent are drugs identified from the pharmacy database documented in the medical record? and (ii) to what extent are indications for prescription recorded. Regarding documentation of drugs, studies have found agreement rates ranging from 89 to 100% between the automated and manual sources.^{29,31,50,51} One of these studies assessed whether indication for a one-time NSAID prescription was recorded in the medical record and found that 7% of the charts ($N = 501$) contained missing or vague diagnoses. There was a correlation between incomplete drug documentation and absence of an indication.⁵¹

With the availability of outpatient diagnoses, it became possible to assess the indication for prescription through automated means. For example, investigators wanted to identify primary care patients who were prescribed antidepressant medications for depression (antidepressants are

Table 15.5. Percent of specific medications obtained from GHC outpatient pharmacies^a

Medication	No. of prescriptions	% always filled at GHC pharmacies
Medicines available only by prescription		
Amitriptyline	66	100
Doxepin	55	100
Ergotamine	35	100
Meclofenamate	61	98
Meprobamate	98	98
Methocarbonol	219	98
Midrin	96	97
Propranolol	38	97
Tolmetin	63	95
Atenolol	31	94
Acetaminophen-codeine	380	93
Diazepam	41	93
Meperidine	53	92
Percocet	116	91
Butalbital-ASA-Caffeine	88	90
Other medicines ^b		
Ibuprofen	869	85
Aspirin	517	30
Acetaminophen	508	24

^aMethods: 2389 GHC enrollees making a visit for back pain, headache, or temporomandibular (TMJ) pain in 1989–90 were asked about all pain medications prescribed or used in the prior month. They were then asked whether each medication was obtained from a GHC pharmacy.

^bThe survey asked about all drugs used in the prior month, not only about drugs prescribed. This could include over-the-counter drugs the subject purchased without a prescription. Thus, for OTC medications, the table indicates the % of OTC use captured in the pharmacy database among all OTC purchases, not the percent among all OTC prescriptions from GHC physicians.

also commonly prescribed for pain and sleep disturbance). They accomplished this by linking the antidepressant prescription record with visit records in the 120 day period preceding the prescription and in the several weeks following the fill, selecting only those prescriptions accompanied by a diagnosis of depression.⁵² This method is far less expensive than chart review.

In a study of antibiotic use at GHC (in collaboration with Harvard Pilgrim Health Plan in Boston, MA), investigators were able to calculate disease-specific antibiotic use rates by linking the pediatric antibiotic prescription to the diagnosis for the most recent ambulatory visit within the preceding three days. For this purpose,

the investigators developed an algorithm to assign a primary diagnosis to each patient when more than one diagnosis existed for a single visit.⁵³

Disease Severity

Because of the high cost of obtaining health status information through survey or medical record review, there has been interest in developing automated measures of health status.⁵⁴ To that end, Center for Health Studies researchers have used the pharmacy database to develop a chronic disease score (CDS) for adults, a weighted sum of medications used for management of significant chronic diseases. There have been several iterations of the chronic disease score over time.⁵⁵⁻⁵⁷ The latest methodology assigns empirically derived weights to almost 30 classes of medication used to treat specific diseases. An individual's CDS is calculated by summing the weights of all classes of medications used (not prescriptions filled) in a 1 year period, in addition to weights assigned to his or her age group and gender, and an intercept term. The CDS predicts future outpatient visit frequency, healthcare costs, hospitalization, and mortality, after controlling for age and gender. Several studies have used the CDS as a proxy measure of health status.⁵⁸⁻⁶⁰ A pediatric chronic disease score has also recently been developed.⁶¹

THE FUTURE

GHC automated data on prescription medicine use and related health care utilization have been used in studies of the effectiveness, adverse effects, utilization, and costs of drugs for more than 20 years. These studies have contributed to our understanding of the strengths and limitations of large automated databases and of the effectiveness and safety of prescription medicines. Future research will focus not only on the effectiveness and costs of the drugs themselves, but also on how medications fit into the overall delivery of health care in ways that improve the effectiveness of care or reduce the costs of providing services. It is of increasing importance that managed health care systems, government, and pharmaceutical

companies conduct research on the cost-effectiveness of prescription medicines and on how medicines can be delivered in the most cost-effective manner possible. In addition to assessing the effects of medicines on biological measures of disease status, the role of medicines in outcomes, particularly improving or harming functioning and overall quality of life, is now recognized as of critical importance. The research programs conducted by investigators of GHC's Center for Health Studies, and by affiliated university-based scientists, are increasingly concerned with these issues of cost-effectiveness and impact on functional status and quality of life.

Additionally, as other HMOs and MCOs develop automated pharmaceutical databases, the future holds opportunity for collaborations across different sites. The resulting large population bases will allow for the study of events possibly associated with drug use, or drug use patterns across many geographic sites. While exciting, broader collaboration carries with it the concomitant challenge of maintaining confidentiality of prescription drug data.

REFERENCES

1. 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.
2. Miller RH, Luft HS. Managed care plans: characteristics, growth, and premium performance. *Annu Rev Public Health* 1994; **15**: 437-59.
3. US Senate Committee on Labor and Human Resources, Subcommittee on Health and Scientific Affairs. *Final Report of the Joint Commission on Prescription Drug Use*. Washington, DC: Government Printing Office, 1980.
4. Wagner EH. Should HMOs do research? *HMO Pract* 1987; **1**: 34-7.
5. Andrews EB, Stergachis A, Hecht JA. Evaluation of alternative methods of assessing pregnancy outcomes using automated indicators of pregnancy [abstract]. *J Clin Res Drug Dev* 1989; **3**: 201.
6. Stergachis A, Scholes D, Heidrich FE, Sherer DM, Holmes KK, Stamm WE. Selective screening for *Chlamydia trachomatis* in a primary care population of women. *Am J Epidemiol* 1993; **138**:143-53.
7. Davis RL, Hasselquist MB, Cardenas V *et al*. Introduction of the new Centers for Disease Control

- Group B Streptococcal Prevention Guideline at a large West Coast Health Maintenance Organization (under review).
8. Wagner E, Curry SJ, Grothaus L, Saunders K, McBride CM. The impact of smoking and quitting on health care utilization. *Arch Intern Med* 1995; **155**: 1789–95.
 9. McBride CM, Lozano P, Curry SJ, Rosner D, Grothaus L. Use of health services by children of smokers and nonsmokers in a Health Maintenance Organization. *Am J Public Health* 1998; **88**: 897–902.
 10. Cherkin D, Deyo R, Battie M, Street J, Barlow W. A comparison of physical therapy, chiropractic manipulation and provision of an education booklet for the treatment of patients with low back pain. *New Engl J Med* 1998; **339**: 1021–9.
 11. Davis RL, Chu S. Antibiotic use patterns and antibiotic-resistant bacteria at Group Health Cooperative, Seattle, Washington. Presented at the HMO Research Network Conference, Minneapolis, MN 1996.
 12. National Cancer Institute. *Surveillance, Epidemiology, and End Results: Incidence and Mortality Data, 1973–1977. (National Cancer Institute Monograph 57)*. Bethesda, MD: National Cancer Institute, 1981.
 13. Taplin SH, Taylor V, Montano D, Chinn R, Urban N. Specialty difference and the ordering of screening mammography by primary care physicians. *J Am Board Fam Pract* 1994; **7**: 375–86.
 14. Taplin SH, Mandelson MT, Anderman C, et al. Mammography diffusion and trends in late stage breast cancer: risk-based guideline implementation in a managed care setting. *Cancer Epidemiol Biomarkers Prev* 1997; **6**: 625–31.
 15. Wagner EH. Population-based management of diabetes care. *Patient Educ Couns* 1995; **26**: 225–30.
 16. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. *Milbank Q* 1996; **74**: 511–44.
 17. Lozano P, Fishman P, Von Korff M, Hecht J. Health care utilization and costs for children with asthma in an HMO. *Pediatrics* 1996; **99**: 757–64.
 18. Simon G, Von Korff M, Barlow W. Health care costs of primary care patients with recognized depression. *Arch Gen Psychiatry* 1995; **52**: 850–6.
 19. Taplin SH, Barlow W, Urban N et al. Stage, age, comorbidity, and direct costs of colon, prostate, and breast cancer care. *J Natl Cancer Inst* 1995; **87**: 417–26.
 20. Chen RT, Glasser JW, Rhodes PH et al. Vaccine Safety Datalink Project: A new tool for improving vaccine safety monitoring in the United States. *Pediatrics* 1997; **99**: 765–73.
 21. Davis RL, Vadheim C, Black S, Shinefield H, Chen R, and the Vaccine Safety Datalink Workgroup. Utility of immunization tracking systems for self assessment, improvement, research and evaluation, and linkage with statewide networks: experience of the CDC Vaccine Safety Datalink sites. *HMO Pract* 1997; **11**: 13–17.
 22. Payne T, Kanvik S, Seward R et al. Development and validation of an immunization tracking system in a large health maintenance organization. *Am J Prev Med* 1993; **9**: 96–100.
 23. 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.
 24. Von Korff M, Dworkin SF, Le Resche L, Kruger A. An epidemiologic comparison of pain complaints. *Pain* 1988; **32**: 173–83.
 25. Curry SJ, McBride CM, Grothaus LC, Louie D, Wagner EH. A randomized trial of self-help materials, personalized feedback and telephone counseling with nonvolunteer smokers. *J Consult-Clin Psychol* 1995; **63**: 1005–14.
 26. 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.
 27. Barlow W, Davis RL, Glasser J et al. Risk for first seizure following DTP or MMR vaccination: Results from the Vaccine Safety Datalink study of children under age seven (under review).
 28. 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. *J Am Med Assoc* 1989; **261**: 2087–94.
 29. Heidrich F, Stergachis A, Gross K. Diuretic drug use and the risk for hip fracture. *Ann Intern Med* 1991; **115**: 1–6.
 30. LeResche L, Saunders K, Von Korff M, Barlow W, Dworkin SF. Use of exogenous hormones and risk of temporomandibular disorder pain. *Pain* 1997; **69**: 153–60.
 31. Stergachis A, Sheingold S, Luce BR, Psaty BM, Revicki DA. Medical care and cost outcomes after pentoxifylline treatment for peripheral arterial disease. *Arch Intern Med* 1992; **152**: 1220–4.
 32. 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.
 33. Unutzer J, Simon GE, Pabiniak C, Bond K, Katon W. Patterns of care for adults with bipolar disorder in a large staff model HMO. *Gen Hosp Psychiatry* in press.
 34. Von Korff M, Barlow W, Cherkin D, Deyo RA. Effects of practice style in managing back pain. *Ann Intern Med* 1994; **121**: 187–95.

35. McCloskey LW, Psaty BM, Koepsell TD, Aagaard GN. Level of blood pressure and the risk of myocardial infarction among treated hypertensives. *Arch Intern Med* 1992; **152**: 513–20.
36. Scholes D, Daling J, Stergachis A, Weiss NS, Wang SP, Grayston JT. Vaginal douching as a risk factor for acute pelvic inflammatory disease. *Obstet Gynecol* 1993; **81**: 601–6.
37. Stergachis A, Shy K, Grothaus L *et al*. Tubal sterilization and the long-term risk of hysterectomy. *J Am Med Assoc* 1990; **264**: 2893–8.
38. Christensen DB, Campbell WH, Madsen S, Hartzema AG, Nudelman PM. Documenting outpatient problem intervention activities of pharmacists in an HMO. *MedCare* 1981; **19**: 104–17.
39. Hartzema AG, Christensen DB. Nonmedical factors associated with the prescribing volume among family practitioners in an HMO. *Med Care* 1983; **21**: 990–1000.
40. Stergachis A, Fors M, Wagner EH, Sims DD, Penna P. Effect of clinical pharmacists on drug prescribing in a primary-care clinic. *Am J Hosp Pharm* 1987; **44**: 525–9.
41. Katon W, Von Korff M, Lin E, Bush T, Ormel J. Adequacy and duration of antidepressant treatment in primary care. *Med Care* 1992; **30**: 67–76.
42. Katon W, Von Korff M, Lin E *et al*. Collaborative management to achieve treatment guidelines: impact on depression in primary care. *J Am Med Assoc* 1995; **273**: 1026–31.
43. Simon G, Lin EHB, Katon W *et al*. Outcomes of “inadequate” antidepressant treatment in primary care. *J Gen Intern Med* 1995; **10**: 663–70.
44. Simon G, Von Korff M, Heiligenstein J *et al*. Fluoxetine vs. tricyclic antidepressants: a randomized trial of effectiveness and cost in primary care. *J Am Med Assoc* 1996; **275**: 1897–1902.
45. Saunders K, Simon G, Bush T, Grothaus L. Assessing the feasibility of using computerized pharmacy refill data to monitor antidepressant treatment on a population basis: A comparison of automated and self-report data. *J Clin Epidemiol* 1998; **51**: 883–90.
46. Davis RL, Marcuse EK, Black S *et al*. MMR2: a comparison of adverse events after MMR given at 4–6 years and 10–12 years of age. *Pediatrics* 1997; **100**: 767–71.
47. Psaty BM, Heckbert SR, Koepsell TD *et al*. The risk of myocardial infarction associated with anti-hypertensive drug therapies. *J Am Med Assoc* 1995; **274**: 620–25.
48. Furberg CD, Psaty BM, Myers JV. Nifedipine: dose-related increase in mortality in patients with coronary heart disease. *Circulation* 1995; **92**: 1326–31 [erratum: *Circulation* 1996; **93**: 1475–6].
49. Kaplan RC, Psaty BM, Kriesel D *et al*. Replacing short-acting nifedipine with alternative medications at a large health maintenance organization. *Am J Hypertens* 1998; **11**: 471–7.
50. Holt V, Daling J, McKnight B, Moore D, Stergachis A, Weiss NS. Functional ovarian cysts in relation to the use of monophasic and triphasic oral contraceptives. *Obstet Gynecol* 1992; **79**: 529–33.
51. 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.
52. Lin E, Katon W, Simon G, *et al*. Low-intensity treatment of depression in primary care: is it problematic? *Gen Hosp Psychiatry* in press.
53. Finkelstein JA, Metlay J, Davis RL, Rifas S, Dowell SF, Platt R. Antimicrobial prescribing for infants and young children. *Arch Pediatr Adolesc Med* in press.
54. Ereth J, Diehr P, Durham M, Hecht J. A tool for administrative decision-making: predicting health status from MIS data. *J Ambulatory Care Manage* 1992; **15**: 30–9.
55. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. *J Clin Epidemiol* 1992; **45**: 197–203.
56. 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.
57. Fishman P, Goodman M, Hornbrook M *et al*. Risk adjustment using automated pharmacy data: the RxRisk Model (under review).
58. Durham M, Beresford S, Diehr P, Grembowski D, Hecht J, Patrick D. Participation of higher users in a randomized trial of medicare reimbursement for preventive services. *Gerontologist* 1991; **31**: 603–6.
59. Wagner EH, LaCroix AZ, Grothaus LC, Hecht J. Responsiveness of health status measures to change among older adults. *J Am Geriatr Soc* 1993; **41**: 241–8.
60. Donahue J, Andrade S, Beck A *et al*. Epidemiologic study of gastric and duodenal safety in patients dispensed alendronate (working paper).
61. Fishman P, Shay D. A pediatric chronic disease score from automated pharmacy data. *Med Care* 1999; **37** (9): 872–80.

16

Kaiser Permanente Medical Care Program: Division of Research, Northern California, and Center for Health Research, Northwest Division

GARY D. FRIEDMAN AND LAUREL A. HABEL

Division of Research, Kaiser Permanente Medical Care Program, Northern California, 3505 Broadway, Oakland, CA
94611, USA

MYDE BOLES

Center for Health Research, Kaiser Permanente Northwest Division, Portland, OR, USA

BENTSON H. McFARLAND

Center for Health Research, Kaiser Permanente Northwest Division, and Oregon Health Sciences University,
Portland, OR, USA

INTRODUCTION

The Kaiser Permanente Medical Care Program, now serving approximately 8.5 million subscribers, is the largest group practice Health Maintenance Organization (HMO) in the United States. The program includes over ten regional divisions. The Northern and Southern California Regions of Kaiser Permanente were recently merged into a single California Division. Investigators in the Kaiser Permanente Medical Care Program, Northern California, have been involved in pharmacoepidemiology for approximately 30 years. The Center for Health Research at Kaiser Permanente

Northwest Division was founded in 1964. A program of drug related health services research began soon after and continues today. In recent years, the availability of an automated ambulatory medical record along with other automated medical data systems, including inpatient and outpatient prescription drug systems, have led the Center for Health Research also to become more deeply involved in pharmacoepidemiology and postmarketing surveillance studies. In this chapter, we review the role of the Kaiser Permanente systems in pharmacoepidemiology research. We begin with a description of the systems, discuss their strengths and weaknesses in conducting

pharmacoepidemiology research, provide examples of work done using these resources, and conclude with some thoughts about where they are going in the future.

DESCRIPTION

KAISER PERMANENTE NORTHERN CALIFORNIA REGION

Clinical Organization

Kaiser Permanente's largest and oldest regional division is in Northern California (initially centered in the San Francisco Bay Area), where the program now serves over 2.8 million subscribers. Approximately 30% of the general population in the geographic areas served belong to the health plan. The region spans a 14-county area that includes the San Francisco Bay and Sacramento metropolitan areas, the northern San Joaquin valley, and Sonoma, Napa, and Fresno counties. Comprehensive health care is provided by Kaiser Permanente medical staff at 15 hospitals and 31 outpatient facilities located throughout the region. Only a small amount of care (e.g., out-of-area emergency visits, radiotherapy) covered by the plan is provided at non-Kaiser Permanente facilities.

Most subscribers and their families join the Kaiser Health Plan as part of an employment group, but about 10% enroll as individuals or families, provided that they pass a health examination. Approximately 12% of members receive some Medicare coverage. Members come from all ethnic and age groups and all occupations, but the indigent and very wealthy are probably under-represented. After the initial year or two of membership, when there is a relatively high turnover, subscribers tend to stay with the program for relatively long periods of time. This makes it possible to use the members' medical records for cohort studies with long followup periods and for case-control studies with long followback intervals from disease diagnosis to exposure.

Each subscriber receives a unique medical record number that is used for all encounters with

the program. This makes it easy to link various records, e.g., pharmacy records with hospitalizations. Computer membership files that contain all current members at a given point in time can provide basic denominator counts for age- and sex-specific prevalence and incidence rates.

The amount of computer stored diagnostic and clinical data has increased substantially over the last 25 years. Since 1971, all hospitalizations in the Northern California program and the diagnoses that resulted from them have been stored annually on computer tapes. This makes it possible to search for hospitalizations of interest among members of a study group, or to identify cases of a particular disease for further investigation. Laboratory and pathology data are now stored in computer databases and radiology/diagnostic imaging data have been computer stored since 1992. Information on outpatient visits was computer stored as of 1994. This outpatient database can help identify individuals with diseases that lack a laboratory diagnostic and rarely result in hospitalization. Despite the growing availability of computerized diagnostic and clinical data, most epidemiologic studies in this setting still require some review of manual medical charts, at least for validation of certain computer stored data.

Northern California Division of Research

The Division of Research was established in 1961, primarily to develop technology based services for the support of medical care. Division researchers have conducted numerous studies over the last 37 years, covering a broad range of topics. The division staff is comprised of individuals with expertise in various disciplines including epidemiology, biostatistics/biometrics, applied behavioral science, economics, data management, research grants and contracts administration, and numerous clinical specialties. In addition, close collaborative ties exist with local universities and departments of health.

The current mission of the Division of Research is to conduct, publish, and disseminate high quality epidemiologic and health services research to improve the health and medical care of Kaiser Permanente members and the society at large. It

seeks to understand the determinants of illness and well-being and to improve the quality and cost-effectiveness of health care.

Pharmacoepidemiology in the Northern California Region

Pharmacoepidemiology first became a major focus of the Northern California Program's Division of Research (formerly the Department of Medical Methods Research) in the late 1960s, as a result of a contract awarded by the US Food and Drug Administration to Dr Morris F. Collen, then director of the division. The aim of the program was to develop a system to monitor adverse drug reactions in both inpatients and outpatients, first in one medical facility and then regionally if feasible. When the contract ended in August 1970, only an outpatient system was sufficiently developed to be operational. It was located in the program's medical center in San Francisco, which then served about 120 000 subscribers. There were two data collection components: a computerized pharmacy system and a clinic diagnosis system. These systems continued to collect data until August 1973, supported by a grant from the National Center for Health Services Research and Development.

The computerized pharmacy system recorded the dispensing of about 78% of the outpatient prescriptions issued by physicians in the facility, as determined by a small followup survey. The pharmacist recorded each prescription dispensed, using a typewriter terminal connected to a central computer system that contained a unique computer medical record for each patient.¹ The data entered included the identifying numbers of the patient and physician, the date and time, the trade name, strength, and form of the drug, the amount dispensed, the number of refills permitted, and the "sig." or instructions to the patient. The terminal printed a label for attachment to each drug container and the pharmacist verified, on-line, the accuracy of the information.

The diagnostic data were recorded by the physicians at the facility. Each specialty clinic had its own form on which were printed the 200–300 diagnoses most commonly made at that clinic,

with provision for writing in the names of conditions not listed. For each condition diagnosed, the physician could indicate whether it was new (including recurrent), old or continuing, or worsening. In addition to writing a progress note, the physician was to mark the appropriate diagnoses on the form, on which a clerk had also recorded the patient's and physician's identifying numbers, date, and time. The information on these forms was entered into the patient-specific computer medical records, first by optical reader, and later by data entry operators using interactive typewriter terminals.

As a result of the 1969–1973 program, a considerable volume of data was collected. Computer stored records include 1 307 767 prescriptions for 3446 drug products dispensed to 149 139 patients. A total of 217 768 patients had diagnoses from one or more clinic visits stored in these records. Most patients' computer records showed a series of intermingled clinic and pharmacy visits that could be used to study the relation of drug use to subsequent clinical events.² A cohort of 143 574 persons with satisfactory identifying information was derived from the 149 139 with records of prescription dispensed.

As of 1996 this cohort had more than 2.18 million person years of followup within the Kaiser Permanente program. A total of 45 625 persons (31.8% of the original 143 574) were still members. In the previous three years, the average annual decrease in membership was about 3800 cohort members per year. It should also be noted that older people are less apt to leave the plan. For example, of persons age 50–59 at entry (defined by their first computer stored pharmacy visit), 33% were still enrolled. Much of the loss was, of course, explained by death. For example, of the persons in this age decade at entry who were no longer members in 1980, 44% had died. Only 19% of the original group had left the Kaiser Permanente program for other reasons.

In 1991, the Northern California region again began computer storing outpatient pharmacy records. The Pharmacy Information Management System (PIMS) was phased in over a three-year period and has been operational at all 108 Kaiser pharmacies since 1994. Data on each prescription

filled are entered into the system in real time by pharmacy personnel before the prescription can be issued. Recorded information includes the patient's and prescribing physician's identification numbers, the drug name, strength, and route; date dispensed; treatment regimen; and days supply. Approximately 90% of the Health Plan membership has a pharmacy copayment of \$10 per prescription per month, or less. However, the results of two recent surveys suggest that approximately 15–20% of adult members fill at least some of their prescriptions at non-Kaiser pharmacies. Since 1994, this system has recorded information on approximately 15 million prescriptions per year.

KAISER PERMANENTE NORTHWEST DIVISION

Clinical Organization

Kaiser Permanente, Northwest Division (KPNW) is a federally qualified, prepaid health maintenance organization that has operated in the Portland, Oregon–Vancouver, Washington metropolitan area since World War II. KPNW provides comprehensive medical care for over 430 000 members in 20 outpatient facilities located in the Portland–Vancouver area and in two smaller communities, Salem, OR, and Longview–Kelso, WA, each about 50 miles away. The demographic and sociographic characteristics of KPNW members correspond very closely to the metropolitan population as a whole (Table 16.1). The membership provides a well defined population base with known dynamic properties.

Most of the medical care is provided in the outpatient setting but KPNW owns and operates one hospital in Portland and contracts for beds in other Portland community hospitals and in Salem, OR, and Longview–Kelso, WA. The 550 physicians of Northwest Permanente, PC, provide the medical care for these members with the exception of a few specialized services that are performed under contract in the community. The terms of insurance coverage by Kaiser Permanente serve to concentrate members' medical care within the walls of KPNW. This is particularly true for

Table 16.1. Demographic characteristics of Kaiser Permanente Northwest Division members (percent of total population^a)

Characteristic	KPNW (members) ^b	Portland–Vancouver area ^c
Age		
0–18	29	27
19–64	59	60
65+	12	13
Gender		
Male	41	49
Female	59	51
Race		
White	94	93
African American	2	2
Asian American	3	2
Native American	1	1
Other	1	2
Hispanic	2	4
Non-Hispanic	98	96

^aNumbers may not add to 100 due to rounding.

^b1995 Membership Survey.

^c1990 US Census data.

chronic diseases. Medical care in non-KPNW settings is largely confined to emergencies and out-of-area events. Almost all elective surgeries are carried out by KPNW surgeons in KPNW facilities. KPNW also operates a prepaid group practice dental care program, which currently enrolls about 190 000 members.

Benefits provided by KPNW include hospital and surgical services, maternity care, x-rays, mammography, laboratory tests, allergy testing, home health care, doctor office visits, well baby care, immunizations, emergency care, mental health and chemical dependency treatment, and routine checkups. More than 90% of the membership has a prepaid drug benefit, and for those without a drug benefit prescriptions are provided at or below prevailing community charges.

Center for Health Research

The Center for Health Research is a professionally autonomous, multidisciplinary research organization that conducts health related research in the public interest, using Kaiser Permanente as its primary laboratory. The center is dedicated to

answering universal and general prevention, epidemiologic, and health services research questions, and this research is supported from a variety of sources, including governmental grants and contracts, private and corporate grants, and a basic support grant from Kaiser Foundation Hospitals.

The center brings experienced university and center based researchers in various disciplines—including biostatistics, dentistry, economics, epidemiology, genetics, gerontology, medical care organization, medicine, nursing, nutrition, pharmacy, psychology, public health, social psychology, and sociology—to a setting where extensive data are available from an operating medical care system that encourages research dedicated to serving the public good through the pursuit of knowledge.

KPNW Databases

The hallmark of the KPNW data systems is the KPNW medical record. Since its inception, KPNW has used unique and permanent health record numbers to identify all members. All encounters with the medical care system—e.g., clinic visit, telephone call, prescription, operation, laboratory test—are linked by this identifier. Kaiser Permanente has retained paper copies of the health records of all members since its founding in 1942. A single, comprehensive medical record for every member of the Health Plan is stored in a central location. Every contact an individual makes with the medical care system is recorded on this chart. Whenever a patient has an appointment at a clinic, his record is delivered to the attending physician to be available at the time of appointment. This record includes hard copy printouts of electronic medical records.

KPNW Administrative Data Systems

The *Membership Information Processing System (MIPS)* maintains the base of membership data and employer group data to support member and patient identification, group contact and benefit administration, group and direct-pay billing, and membership and revenue reporting. The *Kaiser Appointment/Registration/Encounter System (KARE)* provides data on all medical office

visits for each KPNW member. The primary purpose of this administrative system is to assist medical offices in managing appointments.

The *Outside Claims and Referral System (OS-CAR)* is KPNW's automated claims processing system for covered services provided by non-KPNW providers on a fee-for-service basis (some outside services are obtained under capitation contracts); it also identifies covered expense for outpatient drugs purchased at non-KPNW pharmacies.

KPNW Clinical and Dental Data Systems

The *Inpatient Admission/Discharge/Transfer System (ADT)*, implemented in 1980, serves as the basis for the discharge abstract for all patients admitted to a regional KP hospital. This system also includes ambulatory surgical procedures and other major ambulatory procedures performed in the hospital operating suites since 1965. The ADT system also incorporates data from non-Kaiser hospitals.

The *Outpatient Pharmacy System (TOPS)* began in 1986 and records all prescriptions dispensed by KPNW's outpatient pharmacies. Each prescription file contains 25 variables, including quantity dispensed, days supplied, refill number, date, NDC number and other product information. More than two million outpatient prescriptions are dispensed by KPNW pharmacies annually.

The *Automated Inpatient Medication System (RAINS)* is an inpatient pharmacy computer system that captures all inpatient medication orders; it provides a complete history of patient medication orders and stores the history via a unique hospital stay number that is generated on admission.

The *EpicCare* is an automated clinical information system that links data from several data bases to provide clinicians with immediate on-line information necessary for patient care. This computerized medical record has eliminated outpatient paper based clinical reporting and streamlined it into one on-line access point. Periodic write-offs to end-user files will make this rich data source available to researchers.

Adverse and Allergic Drug Event Reporting. Adverse events are entered into a database, and reports are prepared for the local KPNW Formulary and Therapeutics Committee. Data are also sent to the Food and Drug Administration's MedWatch system.

The *Advice Nurse Documentation System (RANDS)* is an on-line system that documents the advice and triage functions of KPNW's telephone advice nurses.

The *Dental Administrative and Clinical Tracking System (TEAM)* provides data on all dental office visits since 1987 in the KP Dental Care program and is updated at each patient visit to include demographic and benefit information, as well as the type of dental service provided.

Emergency Psychiatric Service Logbook. KPNW provides 24-hour emergency mental health services and, each year, the Emergency Psychiatric Service handles about 8000 telephone calls and 2000 face-to-face contacts.

The *Emergency Department Database* contains one line of information for each visit, including date, patient name and health record number, and presenting problem.

The *Laboratory Information System (LIS)* provides an integrated base for the medical laboratory to support processing and reporting of laboratory procedures that have taken place since mid-1992. Major functions of the system include order and specimen tracking, results interpretation and reporting, and cumulative patient reporting.

The *Cytology Database* and *Histology Database* contain the results of all outpatient and inpatient clinical and laboratory procedures starting in August of 1970. More than 85% of the cytology database is Pap smear results. The histology database comprises results on surgical specimens.

The *Radiology Information Management (RIM) Database* contains patient-specific information on all patients treated in any radiology department, including radiology, ultrasound, magnetic resonance imaging (MRI), nuclear medicine, and computerized tomography.

Triple Marker Prenatal Screening Database (MSAFP). In 1991, KPNW began offering to all pregnant women a prenatal screening protocol for

Down syndrome and open neural tube defects commonly known as triple marker screening. Demographic, pregnancy, and screening outcome data are maintained in this database for all women who accept the screening protocol. Pregnancy outcome data are maintained for all women who screen positive for either risk.

Immunization Database. This database is an on-line chart abstraction system that maintains immunization records by individual member. Begun in the early 1980s to provide immunization records for school age children, the database was expanded in 1985 to include all members.

Medicare Plus II Database. The participants in this special program complete a questionnaire annually that includes standardized instruments measuring level of functioning (Activities of Daily Living and Instrumental Activities of Daily Living) and depression (Center for Epidemiologic Studies-Depression scale). This database records information from 1985, with an overall response rate of 90%.

The *Continuing Care Service Database* includes information about diagnosis and treatment for home care services for members who are home-bound.

KPNW Disease Registries

Tumor Registry. KPNW maintains a centralized registry of all cancer patients diagnosed since 1960. Excluding basal cell carcinomas of the skin, more than 35 000 tumors have been registered. The registry tracks the entire course of a patient after a cancer diagnosis. Survival information is acquired through an aggressive followup program that achieves a rate of more than 95%.

The *Benign Breast Disease Registry: Pilot Phase* includes data on nearly 10 000 women diagnosed between 1970 and 1994.

The *Kaiser Permanente Interregional Genetics Registry* contains data on cytogenetic testing on more than five million members of the Northwest Division and the Northern California and Southern California regions from 1986 forward. The Hawaii Region joined the registry in 1992.

The *Breast Cancer Family Registry: Pilot Phase* is funded by a grant to explore the relationship

between benign breast disease and breast cancer, and the genetic changes that lead to breast malignancy. The registry contains nearly 10 000 women diagnosed between 1970 and 1994.

The *Diabetes Registry* is designed to improve the care of KPNW members with diabetes. Patients were identified by reviews of pharmacy records, inpatient and outpatient records, and diabetes clinic logbooks. The registry now includes more than 25 000 persons.

The *Rheumatology Registry*, which began in 1982, logs all rheumatological diagnoses and consultations, identifying members with specific rheumatological diagnoses. As of 1992, the on-line registry contained more than 8000 files.

Center for Health Research Databases

Outpatient Utilization System (OPUS). Since 1966, the CHR has maintained an outpatient utilization database for a continuously updated sample of the KPNW membership. Each month, a simple random sample of all new KPNW subscriber units is added to the pool of enrollees whose records are abstracted; the sample was 5% of Health Plan members until 1984 when it was reduced to 2% for the under-age-65 population only.

Current Membership Study Database. The surveys used for this study query KPNW members in the Northwest Division about their satisfaction with medical services and on their characteristics, such as health status and behaviors. The study surveys a 2% random sample of health plan members each year, with monthly batches of survey questionnaires sent to members—more than 95 000 since the study began in 1975.

Household Interview Survey Database. In 1970, the CHR initiated an intensive household interview survey of a subset of the OPUS population. The survey elicited extensive individual and family information, including objective, factual information such as demographic and socioeconomic characteristics of the families and individuals in the units, information about their health status and behaviors, perceptual or attitudinal information, including beliefs and opinions about health

and medical care, and other social and psychosocial data.

The *Common Recruitment Pool (CRP)* is a computerized database system designed to hold information about individuals who could become volunteers or may have already volunteered for screening in a CHR clinical trial.

The *Common Control Pool (CCP)* contains the basic demographic and eligibility data for virtually all people who have ever been members of the Kaiser Foundation Health Plan, NW Division.

The *Clinic Appointment System (CAS)* is a computerized appointment database system designed specifically for use in randomized clinical trials. A variety of standard reports are produced from CAS and *ad hoc* queries may be issued, resolved, and reported if appropriate.

Medical Visit Survey Database. Since 1990, the CHR has conducted an ongoing survey of a sample of KPNW members, stratified by medical office, who have made recent visits to KPNW clinicians. Survey questionnaires are sent shortly after the visit, and focus on satisfaction with the visit.

Dental Visit Survey Database. A random sample survey of members making visits to KPNW dental clinics has been conducted since 1994.

Social Health Maintenance Organization (Social/HMO) Database. The Social/HMO is a national demonstration at four sites designed to evaluate the ability of a single delivery system—the HMO—to provide elderly members with a comprehensive package of preventive, acute, and long term services.

Vital Statistics Database. The CHR has access to various vital statistics maintained by the states of Oregon and Washington, including records of live births from 1965 and deaths from 1975. This database includes infant mortality from 1986 to 1991 and records of therapeutic abortions from 1975 to 1983.

Pregnancy Registry. The registry uses diagnoses, laboratory data, obstetric clinic visits, and obstetric ultrasound reports to identify pregnant KPNW members. Pregnancy outcomes of spontaneous abortion, therapeutic abortion, live birth, and still birth also are included in the registry.

*Pharmacy Database in Pharmacoepidemiology
Research*

The KPNW pharmacy database has been in operation since 1986. Table 16.2 lists the top 50 most frequently dispensed drugs and number of users in KPNW in 1997.

Automated pharmacy data can be employed to identify users of drugs. Automated medical records can also be searched to identify individuals with particular diagnoses who can then be followed. All data can be linked through unique identifiers. Cohort members can be followed as long as they remain enrolled in the health plan. Person time at risk can be calculated easily from the enrollment data. CHR researchers have conducted a number of long term followup studies on cohorts of medication users.

Some out-of-plan use of pharmaceuticals may not be recorded in the database; however, over 90% of the KP membership has some type of drug benefit that makes it financially advantageous to fill prescriptions at a KPNW pharmacy. Also, KPNW drug prices are generally equal to or lower than those at outside pharmacies. Finally, surveys of KPNW membership show that the vast majority of members fill their prescriptions at a KPNW pharmacy.

STRENGTHS

The major strengths of the Kaiser Permanente Medical Care Program as a site for pharmacoepidemiology studies are mostly apparent. One is the size of the population available—over 2.8 million in the Northern California Region. KPNW currently has more than 430 000 members, representing about 25% of the population of the Portland, OR and Vancouver, WA, metropolitan area.

A second advantage is the representativeness of the population served by Kaiser Permanente. Although there is a relative lack of the poor or the very wealthy, the program is much more representative than a number of the alternative data systems.

Table 16.2. Top 50 Prescription Medications and Exposure for 1997

Medication	Users
Amoxicillin	129 614
Ibuprofen	102 852
Acetaminophen with hydrocodone	86 578
Trimethoprim/sulfamethoxazole	77 136
Albuterol	60 982
Beclomethasone dispropionate	60 542
Acetaminophen with codeine	58 056
Cimetidine	49 304
Estrogens, conjugated	44 581
Triamcinolone acetonide	39 088
Erythromycin	37 105
Codeine	36 626
Lisinopril	35 565
Cephalexin	31 939
Atenolol	28 914
Medroxyprogesterone	25 889
Hydrocortisone	25 258
Hydrochlorothiazide	24 709
Prednisone	24 210
Penicillin V potassium	22 799
Levothyroxine	20 164
Fluoxetine	19 628
Potassium chloride	18 615
Nitroglycerin	18 470
Paroxetine	17 139
Methocarbamol	17 001
Ciprofloxacin	16 941
Phenylpropanolamine	16 701
Naproxen	16 271
Amoxicillin/clavulanate	15 819
Neomycin/polymyxin	15 500
Metronidazole	13 095
Amoxicillin trihydrate	11 790
Doxycycline	11 691
Sodium sulfacetamide	11 375
Cyclobenzaprine	11 301
Oxycodone/acetaminophen	11 289
Fexofenadine	10 767
Omeprazole	10 426
Trazodone	10 291
Furosemide	10 015
Norethindrone/ethinyl estradiol	10 007
Azithromycin	9 334
Glyburide	9 049
Lovastatin	9 026
Amitriptyline	8 966
Triamterene/hydrochlorothiazide	8 966
Hydroxyzine	8 821
Sodium fluoride	7 935
Benzonatate	7 875

A third advantage is the racial and ethnic diversity of the membership in Northern California, which is approximately 65% Non-Hispanic White, 10% Hispanic, 7% African-American, 14% Asian-American, and 3% other. In contrast, although the KPNW membership mirrors the characteristics of the population in the geographic area it serves, it is strongly represented by white, middle class, employed individuals.

Fourth, the stability of the health plan population after the first year of membership permits followup of patients for long term drug effects. In KPNW, the median length of enrollment retention exceeds five years. For 0–14 year olds, the five-year retention percentage is 63%. Retention is lower for younger adults and higher for older adults. The Kaiser Permanente membership is not subject to the turnover inherent in programs in which eligibility changes with income.

Fifth, the existence of membership files allows one to enumerate those at risk for a drug exposure or disease and to calculate true prevalence and incidence rates.

Sixth, the use of a unique medical record number for each member allows the linkage of drug dispensing with inpatient and outpatient diagnosis data.

Finally, and probably most important, access to primary medical records is reasonably complete and fairly easy. This permits validation of diagnostic information and access to information on confounding variables that is not available in the computerized record.

WEAKNESSES

Like any other data system, the use of Kaiser Permanente data for pharmacoepidemiology studies also has its disadvantages. Probably foremost among these in the past, before the development of PIMS, was the absence of comprehensive computerized pharmacy information in Northern California, with the exception of the computerized pharmacy systems that were in effect for short periods of time.

A second disadvantage was the absence of confirmed computerized outpatient diagnostic data in Northern California before 1994. While automated systems to capture diagnostic information on outpatient visits now have been implemented, this information is sometimes incomplete and often needs to be confirmed with data from other sources, such as the manual medical record.

Third, there is limited demographic information on the Kaiser membership. Age and gender are collected on the full membership. Race/ethnicity is only available on hospitalized individuals and subgroups who have completed special questionnaires, and there is no system-wide collection of information on marital status, education, income, or occupation.

Fourth, although the Kaiser databases are a rich source of patient risk factors and confounders, not all relevant variables will be present in the automated data. For example, smoking status and alcohol consumption are not routinely recorded in the computerized databases. Alternative methods of collecting these data, such as survey or medical record review, will be required.

Finally, as with other HMOs, Kaiser Permanente formularies are somewhat limited. Many of the newest and/or most expensive drugs may not be available for study. It is not uncommon for only one brand of a particular drug to be available on the formulary at a point in time. This prevents investigation of head-to-head comparisons of drugs for effectiveness and toxicity. Nonlisted products prescribed by individual physicians are usually stocked. However, one cannot adequately study a drug that is rarely dispensed.

PARTICULAR APPLICATIONS

Kaiser Permanente data are particularly useful for pharmacoepidemiology research when one needs a large population, especially a representative population, and if one needs prolonged followup. Kaiser Permanente data are easier to retrieve when one is studying a disease that results in hospitalization, especially when major confounding variables

also result in hospitalization or, at least, are recorded in the medical record. The use of Kaiser Permanente data is problematic if a population larger than a few million is needed, or if information is needed which is not usually available in a medical record.

In general, cohort studies using Kaiser Permanente data can be performed by using computerized pharmacy data to identify drug exposed individuals and computerized and/or medical record diagnosis data to evaluate outcomes. Case-control studies using Kaiser Permanente data can be performed by using computerized and/or medical record diagnosis data to identify cases, membership files to identify controls, and then computerized and/or medical record pharmacy or drug prescribing data to evaluate exposures. Examples of each will be provided.

NORTHERN CALIFORNIA KAISER PHARMACOEPIDEMOLOGY STUDIES

Initial Phase

Our initial efforts at the analysis of the 1969–1973 data consisted primarily of a search for associations between drugs and subsequent clinical events. Incidence rates in users of a drug or group of drugs were compared with those in nonusers. At that time we were primarily concerned with short term drug effects. Using the data collected during the first six months of computerized pharmacy records, it could be shown, for example, that yeast infections of the vagina were about twice as common among users of oral contraceptives as among nonusers in the same age range, and that these infections occurred much more commonly among users of the “combination” than the “sequential” oral contraceptives.³

After data collection ended in 1973, we conducted a few additional studies with the accumulated data, including analyses of drug use that employed the pharmacy data and analyses of disease frequency that used the clinic data. We also carried out one study of an adverse drug reaction that provided some insight into the limitations of the computer stored diagnostic data. We investigated the incidence of diarrhea following the use of

the antibiotic, clindamycin.⁴ Instead of relying solely on the computer stored diagnoses to detect the occurrence of diarrhea, we examined the medical charts of all clindamycin users identified in the computer records. This experience confirmed our suspicion that the computer stored diagnoses relating to this event were incomplete. In eight weeks of followup of some 300 persons who received clindamycin, there were ten cases of diarrhea recorded in the medical charts but only two cases recorded in the computer records. Despite the under-recording of diarrhea in the computer records, a comparison of the incidence of diarrhea among clindamycin users and nonusers showed a 6.5-fold greater occurrence in users, which was statistically significant ($p < 0.05$). Thus, the computer data alone could still be used to detect this adverse reaction, despite their demonstrated incompleteness.

There were two main reasons why some of the computer records did not show this diagnosis. First, some of the clinics prescribing clindamycin (e.g., dermatology and ear, nose, and throat) did not have diarrhea listed on their data recording form, since it rarely comes to the attention of physicians practicing these specialties. Second, on other patients it seemed that some physicians were satisfied to record this new symptom in their notes and did not feel obligated to list it as a diagnosis. Thus, we learned that information stored by selecting from a limited list of conditions in computer medical record systems will not necessarily reflect all the problems a patient has. This is an extension of a limitation that is inherent also in conventional medical charts—that the busy physician does not have time to record every detail reported by the patient.

Surveillance for Possible Carcinogenic Effects using 1969–1973 Pharmacy Data

Study Design

Using the 1969–1973 computer stored data, we developed a research program to detect carcinogenic effects of drugs. The problem of medicinal drugs as actual and potential causes of cancer has been reviewed by Selby *et al.*⁵ The list of drugs

already shown to have carcinogenic effects includes radioisotopes, immunosuppressive drugs, cytotoxic drugs, drugs that contain arsenic, phenacetin, and coal-tar, and certain hormones. Suspicion has been raised about a variety of others based on case reports and epidemiologic studies. Also, laboratory research has pointed out substances that are carcinogenic for animals and mechanisms of carcinogenesis that throw suspicion on certain drugs used by humans. Given the widespread use of drugs, the suspicions that have already been raised, and the seriousness of the adverse reactions under consideration, the importance of further work in this area is obvious.

In 1977 we initiated a two phase program to study drug carcinogenesis (currently supported by grant No. R35 CA49761 from the National Cancer Institute). Through 31 December, 1996, 15 533 incident cancers had been documented in the 1969–1973 San Francisco pharmacy cohort. At that time, the number of cancers was growing by about 500 per year. The hypothesis seeking phase consists of followup of the 143 574 persons with computerized pharmacy records and comparison of the incidence of all types of cancer among users of various drugs with that among nonusers. The hypothesis testing phase involves mainly detailed case–control studies of Northern California Kaiser Permanente patients using their longitudinal medical records, but also other data analyses and study designs. The hypotheses to be tested are selected from those developed from the hypothesis seeking activities, as well as others appearing in the medical literature.

Among the 3446 drug products dispensed during the four years, there were 95 drugs given to at least 1000 people. Some drugs of interest had very large numbers of recipients. Tetracycline, one of the most commonly recorded drugs, was given to about 23 000 patients; it merits study because it is a tertiary amine with the potential for conversion to carcinogenic nitrosamines in the stomach if nitrites are also present. Another very commonly dispensed drug was promethazine. Over 17 000 patients received it in expectorant mixtures. This drug is also of interest since the phenothiazines are said to be stimulators of prolactin release and prolactin may be positively related to breast cancer development.

We have used two sources of data to detect subsequent cancer occurrence. Both of these can be linked to the drug dispensing data by means of each patient's unique medical record number which follows her or him through the Kaiser Permanente system. One source is the computerized hospitalization records for the entire Northern California region. The discharge diagnoses are coded by medical records personnel and included with each hospitalization record. These cover the years 1971 to the present. The other major source of followup is the cancer registry of the Surveillance Epidemiology and End Results (SEER) program now operated by the Northern California Cancer Center. This initially covered those Kaiser Permanente hospitals located in the five counties closest to San Francisco. These records go back through 1973 and records from the same hospitals were obtained from the Third National Cancer Survey, covering the years 1969–1971. This cancer surveillance has been extended throughout the state, as now required by California law, and tumor registry coverage in Northern California Kaiser Permanente has been complete since 1990. In order to exclude from followup persons who already had cancer when followup began, those cases who were hospitalized in the San Francisco facility in 1968 were added to the cancer records, by manually reviewing listings in that hospital's record room.

Our routine followup coverage does not extend beyond the Kaiser Permanente system, because losses to followup are small. After members have been in the Health Plan for two or more years, the dropout rate averages 3% per year. Also, because of the costs of cancer care, there is strong motivation for the cancer patient to continue his or her care within the Kaiser Permanente prepaid system.

We tested the adequacy of the within-Kaiser Permanente followup by submitting a sample of 10 252 subjects from the pharmacy cohort (every 14th record of the 143 531 subjects with names recorded) to the SEER program. Altogether, 437 cancers were found by SEER to have developed in this group by the end of 1982. Of these, 66 or 15.1% were clearly missing from our files. (Similar cancer codes in our files were not counted as

missing.) Of these, 17 were diagnosed while the subject was a member and 49 while not a member of our Health Plan. Thus, altogether about 15% of cancers in 9–13 years of followup were known to be missed. Some additional cancers probably developed outside of both Kaiser Permanente and SEER surveillance, and the Kaiser Permanente data contain cancers not recorded by SEER. Our surveillance for cancer within Kaiser Permanente records for the duration of membership appeared quite complete, in that only 17 cancers were diagnosed outside of Kaiser Permanente during membership and missed by our surveillance. This represents 17/388 or 4.4% of those cancers that occurred during membership according to SEER records.

We estimated the effect on the minimal detectable relative risk of having 18% (100%/ [100% – 15% missed]) more cancers detected in the screening analysis than are now detected with our Kaiser-only followup. Our calculations were based on a Poisson distribution, which is used in the relevant statistical significance testing. The error (significance) level was set at 0.05 (two-sided) and power at 0.80. The results appear in Table 16.3 for a wide range of expected numbers of cancers. Clearly the addition of 18% more cancers leads to little improvement in the sensitivity of the study.

Table 16.3. Effect of Adding Out-of-Kaiser Permanente Surveillance for Cancer Cases on the Minimum Detectable Relative Risk in Drug–Cancer Studies^a

Expected number of cancers	Current approach	If 18% more cancers
0.5	8.60	7.29
1	5.50	5.68 ^b
2	3.95	3.86
5	2.74	2.51
10	2.15	2.01
20	1.72	1.69
40	1.50	1.45
80	1.34	1.31

^a We thank Bruce Fireman, MA, for assistance.

^b The anomalous result of a higher minimum detectable relative risk with 18% more cancers when there is one expected case is due to the fact that changing the expected from 1.00 to 1.18 and requiring that the observed number to be an integer changes the observed number required for significance from four to five.

The search for additional cancer cases in SEER records proved to be quite costly and, as seen, yielded little benefit to the study. Thus, we decided not to expand the surveillance in this way routinely. However, where there is an especially interesting or important association, or lack of association, observed in our data or elsewhere, a special *ad hoc* search for additional cases in users of a particular drug may be worthwhile.

Method of Analysis

We use traditional analytic methods to screen for possible carcinogenic effects. Users of the drug or drug group to be evaluated are compared with the entire pharmacy user cohort listed in the 1969–1973 computer records with respect to the incidence during the ensuing years of each major type of cancer and of any cancer. Followup of each user begins at the first computer recorded dispensing of the drug under study. Followup for the entire cohort begins with their first recorded receipt of any drug. Followup ends when the patient leaves the Northern California Kaiser Permanente Program. Of course, after a cancer is diagnosed in a subject, either before or after followup begins, he or she is not considered at risk of developing that cancer.

The users and nonusers are divided into males and females and classified initially according to their age when followup begins and reclassified as they progress into subsequent 10 year age groups. Incidence rates for each cancer and all cancers combined are calculated for each age–sex subgroup on a person year basis. These rates are compared with those for the entire cohort by means of the morbidity ratio, which is merely the drug user incidence rate divided by the cohort rate. An age–sex standardized morbidity ratio is calculated for all drug users. The degree of departure of the morbidity ratio from unity, that is, the magnitude of the association of the drug with the cancer, is evaluated for statistical significance using the Poisson distribution. Naturally, unless a drug–cancer association is hypothesized in advance, each probability (*p*) value obtained may be viewed as an underestimate of the true

probability, since we are looking, in an exploratory fashion, at so many drugs and cancers.

Our comparison of users of a particular drug with all pharmacy users has a bias toward not finding a drug–cancer association since the drug users form part of the entire group. However, users of most drugs comprise such a small fraction of the entire cohort that the bias will usually be very small. Even with one of the largest groups, the 23 000 who took tetracycline, a twofold increase in risk of cancer compared to nonusers would only be reduced to a 1.7-fold increase compared to all pharmacy users. The overriding advantage of using all pharmacy users as our standard of comparison was that the age–sex-specific incidence rates for each cancer had to be computed only once; they are stored and compared with those for each drug user group. If the analytic program had to compute a new set of incidence rates for nonusers every time we studied a different drug, this would be very costly and inefficient.

The Question of Cancer Incubation Periods

After exposure to a carcinogen, the incubation period for cancer is often many years, even decades. Thus, the question arises as to whether we have, in past years, been conducting these analyses prematurely. As of 1996, the over 45 000 persons still in the program had from 23 to 27 years of followup. It is true that a number of known occupational cancers do seem to require 20 years or more incubation from the onset of exposure. However, there are some drug related cancers that suggest the possibility of a much shorter incubation period. For example, the excess risk of lymphoma in relation to immunosuppressive drugs for renal transplant patients appears within a year of their use. Leukemia develops after alkylating agents are given to patients with multiple myeloma after an average latent period of 4½ years.⁶ An example of a shorter incubation period is radiation induced leukemia in adults, where the usual incubation period is 5 years, and 90% of all cases occur within 10 years. An important point is that, although the first indication of drug use would have been recorded in these data in 1969, some patients had used the drugs for quite some

time beforehand, thus lengthening the actual incubation time. Thus, we believe that our analyses to date have been worthwhile. Nevertheless, our chances of finding drug carcinogenesis will increase as followup time increases, provided that not too many subjects are lost to observation. Prior use of a drug may have occurred in some persons regarded as nonusers of that drug. However, for nonusers of most drugs the proportion would have been very small and the effect on our findings negligible.

Hypothesis Seeking by Following Up a Cohort

Why do we consider these analyses primarily to be hypothesis generating rather than hypothesis testing, since cohort followup studies are often considered to be more definitive than case–control studies? The problem is that these computerized data lack the additional information on confounding variables that are so necessary to test hypotheses. Thus, for example, while they contain information on age and sex, they lack information on reproductive history that one would like to have for a study of breast cancer. Also, because they only cover a limited time, they do not provide a full picture of the duration of drug use and the total amount of the drug taken. For these reasons we feel it preferable to test hypotheses by means of case–control studies that involve a careful review of the patients' complete longitudinal medical records. Another possibility that may be appropriate for hypothesis testing in certain situations is to use the cohort type of study and, in addition, investigate comparability with regard to possible confounding factors by examining the charts of a sample of users and nonusers.

Examples of Results of Screening

Our screening analyses have been summarized in four reports to date.^{7–10} One positive finding of great interest has been an association of barbiturate use with subsequent lung cancer.¹¹ Phenobarbital is a known cancer promoter in experimental animals and further analyses of our computer stored data suggested a dose–response effect and that increased cigarette smoking among barbitu-

rate users was not the explanation. However, medical record review failed to confirm dose-response and revealed no restriction of the association with any specific histological type of lung cancer. Also, there were too few lung cancer cases among nonsmoking barbiturate users to be sure that uncontrolled confounding by cigarette smoking was not involved. Sixteen additional years of followup failed to demonstrate that the association was due to confounding by cigarette smoking.¹² We continue to pay close attention to this association as more followup data accumulate. A second finding of interest was the negative association between bladder cancer and phenobarbital use. This has been observed in a few other studies and it has been hypothesized that phenobarbital may induce drug metabolizing enzymes that detoxify bladder carcinogens found in cigarette smoke.¹³ We linked our drug screening data with computerized smoking information that was available on a subset of our cohort because they had received a multiphasic health checkup. Although our estimates were imprecise, our results supported this hypothesis, as we found that there was a negative association between barbiturate treatment and bladder cancer risk only among current and former cigarette smokers.¹⁴

We have also had several negative findings that have been of particular interest in our drug-cancer analyses. When questions were raised about metronidazole as a cause of cancer¹⁵ we showed that, in up to 11 years of followup, there was no increased risk of cancer of any site except uterine cervix, whose association with metronidazole use was explainable on other grounds.¹⁶ Similarly, when it was suggested that use of digitalis was protective against breast cancer,¹⁷ we could find no confirmation in our data.¹⁸ We also tested the hypothesis that conflicting findings regarding rauwolfia and breast cancer were explainable if this drug increased risk only in women over age 50 who took it for at least five years.¹⁹ We could not confirm this hypothesis.²⁰

After looking at biennial analyses of the 215 most commonly used drugs, with followup now extending through 1996, we have been rather impressed at the apparent safety of most drugs, as they are ordinarily used, with regard to cancer

causing effects. Drugs, after all, are "chemicals" that are introduced into the body in relatively high doses compared to the ingested quantity of many chemicals in the environment. However many of our analyses have low statistical power to detect associations. More assurance will come as cancer cases now accumulate after two decades of followup, since the latency period for some environmental cancers can be this long.

Mortality Followup

We also added mortality followup to our screening capabilities. A computer tape file listing 63 486 cohort members no longer in the Health Plan was given to Dr. Max Arellano for CAMLIS²¹ computer matching with State of California death records. This program identifies probable and possible matches between persons in two sets of records with an indication of the strength of the probability that the matches are correct. Our chart reviewers manually checked the identity in our Health Plan records of all matched decedents about whom there was any doubt concerning their membership in the cohort. Altogether 9771 deaths were confirmed through 1980.

We performed selected drug-cancer screening analyses on the mortality data analogous to what was done with the incidence data. We also added total mortality and a few grouped noncancer causes of death to the analysis. The findings for cancer mortality were generally similar to those for cancer incidence. Noncancer findings were also as expected (e.g., tranquilizers and other psychoactive drugs were associated with suicide and other traumatic death).

We have not continued routine mortality followup in our cancer surveillance because it adds relatively little to what we can learn from incidence data.

Research Using Recently Implemented Pharmacy Information Management System (PIMS)

As with the 1969-73 pharmacy data, information stored in PIMS can be used to examine adverse or beneficial outcomes associated with the use of

specific medications. PIMS data can be linked with automated membership records and a variety of clinical and diagnostic data to obtain information on other exposures as well as outcomes. Given the recency of implementation, this database is too young to look at drug–disease associations with a latency period of more than a couple of years (e.g., most drug–cancer associations). However, the short term effect of drugs can be evaluated. Examples of studies using PIMS include examinations of glycemic control among type 2 diabetics among users of hormone replacement therapy²² or metformin,²³ and vaginal bleeding and associated gynecologic care in postmenopausal women using hormone replacement therapy.²⁴

Several studies have used PIMS to examine patterns of drug utilization or drug costs among the Northern California membership at large, or among subgroups based on disease status. Some of these studies have been conducted with a purchased cost accounting software package that integrates the regional utilization databases with the general ledger to provide fully allocated costs by department, medical center, patient, or service. Examples of studies of drug cost include a cost analysis of a *Helicobacter pylori* eradication strategy²⁵ and an examination of health care costs associated with obesity.²⁶ Utilization studies include a comparison of continuation of transdermal versus oral estrogen postmenopausal hormone replacement therapy.²⁷

Another valuable use of our computerized pharmacy data is to help identify individuals with a particular disease condition or disease subgroup. PIMS was an instrumental case finding component of an effort to establish a Northern California Kaiser Permanente Diabetes Registry. This registry includes approximately 100 000 members with type 1 or type 2 diabetes and approximately 80% in any given year were identified, at least in part, from prescriptions in the pharmacy database. PIMS was also part of a computer based model to identify high risk children with asthma.²⁸

Case-Control Studies

A number of case–control studies of adverse drug effects have been conducted over the years using

drug exposure data that have been both computer stored and manually abstracted from medical charts.^{19,29–36} In some studies, use of drugs such as oral contraceptives was assessed primarily by personal interview.^{34–36} Because of occasional errors in automated data and nonuniform diagnostic practices and criteria among the many physicians in our large health care system, the disease under study needs to be verified by medical record review when case subjects are identified. Control subjects representing the overall subscriber population are usually selected from membership rolls, applying matching criteria, as appropriate. Thus, in a case–control study focusing on prior drug use, it is often desirable to match control subjects to cases for age and for length of membership in the health plan.

APPLICATIONS OF KPNW

Regulatory Questions

Table 16.4 provides examples of projects that used KPNW databases to address issues of concern to regulators. These projects have had an impact on national policy. For example, in late May of 1996 the Food and Drug Administration was considering the approval of the antiviral drug sorivudine. Previous work in Japan indicated that sorivudine apparently blocks metabolism of 5-fluorouracil by inhibiting dihydropyrimidine dehydrogenase. Fatalities were reported due to 5-fluorouracil toxicity in users of both drugs. The regulatory question was the extent to which antivirals are used concomitantly with 5-fluorouracil. The KPNW automated pharmacy data for 1995 yielded 5627 acyclovir users and 209 injectable 5-fluorouracil users, of whom seven had been prescribed both medications. Records were sought and retrieved for all seven patients who may have been concomitant users. Record review showed there were four who had been using both drugs simultaneously.

The KPNW research team has also worked with the pharmaceutical industry on postmarketing surveillance. For example, KPNW participated with industry scientists in developing a registry for adverse events associated with use of the antiviral

Table 16.4. Examples of reports to the Food and Drug Administration from the KPNW Center for Health Research

Subject	Date
Sildenafil use	1 November 1998
Fenfluramine/phentermine use	1 July 1998
Ticlopidine use	30 May 1998
Sumatriptan use	4 January 1997
Angiotensin converting enzyme inhibitors and neutropenia	5 September 1996
Ceftriaxone, cholelithiasis, and nephrolithiasis	30 August 1996
Fluoruracil and acyclovir use	29 May 1996
Tamoxifen and macular edema	23 May 1996
Tramadol and seizures	20 May 1996
Alendronate and esophagitis	15 May 1996
Lamotrigene and Stevens–Johnson syndrome	21 April 1996
New molecular entities	15 April 1996
Angiotensin converting enzyme inhibitors and hypoglycemia	15 October 1995
Medication use in pregnancy	1 October 1995
Calcium channel blocker use	21 September 1995
Metformin use	21 September 1995
Norplant use	16 September 1995
Methylphenidate and clonidine use	16 September 1995
Triazolam use	30 June 1995
Tumor promotion and use of antihistamines or antidepressants	28 June 1995

drug acyclovir.⁴⁰ This project involved collaboration among several sites within Kaiser Permanente.

Use and Effects of Psychotropic Drugs

Supported by the Food and Drug Administration, the National Institute of Mental Health, and the pharmaceutical industry, KPNW investigators have examined use of antipsychotic medications,⁴¹ benzodiazepines,^{42,44} lithium,⁴³ and antidepressant drugs^{45–47} in the defined population of KPNW members. These projects have provided a foundation for work on adverse events.

For example, KPNW researchers worked with scientists at the Food and Drug Administration and the National Cancer Institute to examine the contention that antidepressant and/or antihistaminic drugs may promote tumor growth.⁵² The

study was stimulated by several laboratory studies suggesting that the growth of tumors in rodent cancer models could be markedly accelerated by commonly used antihistamine or antidepressant medications at usual doses. The laboratory data were reported prominently in the lay press. The KPNW subjects were identified from the automated tumor registry and included individuals with breast cancer, colon cancer, or melanoma. The cases were defined as patients who had recurrence of their original tumor. Controls were cancer patients with no recurrence. Exposure to antidepressant or antihistamine drugs (during the time at risk following treatment for the original tumor) was determined from the automated pharmacy system. The KPNW tumor registry provided stratification variables such as tumor stage at original diagnosis. As is discussed in detail by Weiss *et al.*,⁵¹ no evidence was found to support the contention that antidepressant or antihistaminic drugs promote tumors.

The rise of managed care has stimulated considerable interest in the new discipline of pharmacoeconomics. Workers at KPNW have addressed these issues by examining antidepressants since these medications have become one of the most expensive therapeutic drug classes.^{45,46} Especially noteworthy in this work is the finding that the overwhelming majority of prescriptions for antidepressant drugs are written by primary care providers.

Drug Use in Pregnancy

Work is under way at KPNW to expand a pregnancy registry derived from automated appointment, diagnostic, and laboratory data systems. The idea is to use automated data to identify pregnancy. The KPNW researchers are interested in pregnancies leading to spontaneous or therapeutic abortion, as well as those leading to live birth or still birth. The pregnancy registry will have several uses, such as examining the impact on the fetus of new anticonvulsant or antidepressant medications. Work to date on the pregnancy registry project has involved validating data from automated systems. Automated pregnancy markers include urine or serum human chorionic

gonadotrophin as well as obstetric ultrasound (sonogram) reports, in addition to outpatient diagnoses and referrals to the obstetrics department. The validation study used the paper medical record as the "gold standard." The validation also allowed the researchers to compare the date of the last menstrual period with the date of the first automated marker of pregnancy. In addition, the validation study allowed the researchers to compare spontaneous abortion reports in the paper record with those in the automated databases. Generally speaking, the automated data appeared to reflect accurately the pregnancy related events described in the paper medical record. There are also linkages between KPNW's deliveries and the Oregon and Washington birth certificate files. Mothers and babies have also been linked. The pregnancy registry can be expanded to incorporate other Kaiser Permanente programs nationally, which have over 50 000 pregnancies annually. Early reports on medication usage during pregnancy have been presented.⁵⁰

Studies Of Biologics and Devices

Both the KPNW Center for Health Research in Portland and the Group Health Center for Health Studies in Seattle participate in the Vaccine Safety Datalink project supported by the Centers for Disease Control and Prevention.^{37,39} This national study involves preparing customized data sets that are transmitted to the Centers for Disease Control and Prevention in Atlanta. The project has also led to theoretical epidemiologic work on methodology for monitoring vaccine safety. As part of the vaccine project, Kaiser Permanente researchers prepare and ship to the Centers for Disease Control and Prevention records from automated data sources pertaining to children. The data include information on vaccinations, enrollment, outcomes (e.g., hospital discharge, emergency department use, or death), and covariates (e.g., birth certificates). Associations inferred from screening analyses are validated by nested case-crossover or case-control studies. Medical records are abstracted to supply data for the detailed studies. Outcome measures of interest include seizures, ataxia, aseptic meningitis, and mortality.

Regarding devices, at the request of Food and Drug Administration scientists, the KPNW research team examined the possibility of studying adverse effects of contact lenses. Studies of this nature appear feasible as KPNW dispenses a wide variety of contact lenses and has a system that records the brand dispensed to a given KPNW enrollee. Roughly 40% of KP members have optical benefits. Also, KPNW researchers have worked with the staff at the Casey Eye Institute's National Registry of Drug Induced Ocular Side Effects, which is located at the Oregon Health Sciences University in Portland.

Similarly, KPNW staff examined (at Food and Drug Administration request) the feasibility of examining the safety of amniocentesis and chorionic villus sampling. The KPNW Center for Health Research maintains a genetics registry that includes data from Kaiser Permanente programs in California, Hawaii, Oregon, and Washington. The genetics registry can be used for studies on devices used in prenatal diagnosis.

Methodologic Work

An important aspect of work at KPNW has been the development of methodology for use in pharmacoepidemiologic and pharmaco-economic studies. For example, KPNW researchers devised procedures for matching health maintenance organization data with state vital statistics (e.g., birth and death) records.³⁸ These techniques are used to examine mortality within the defined population. The KPNW researchers developed a programming method to link Oregon and Washington death certificates to records of Kaiser Permanente members. The method was developed in collaboration with researchers at the Boston Collaborative Drug Surveillance Program and was implemented using advanced SAS programming techniques. This methodology has been used to ascertain infant mortality among the 18 000 newborns delivered in KPNW hospitals between 1986 and 1989 in a study pertaining to the alleged association between fluoride drops and Sudden Infant Death Syndrome.

A key issue for both pharmacoepidemiology and pharmaco-economics is the prevalence of

medication usage. This topic is not straightforward since the underlying health maintenance organization population (while it is well defined) is constantly changing. McFarland has devised statistical techniques that can be used to compare period prevalence derived from administrative databases with survey data and/or with one another.⁴⁸ The statistical methodology deals with the varying times under observation.

Another issue of considerable interest is the dose of prescribed medications. KPNW workers have devised methodology that can be used to translate automated pharmacy system dispensing data into average daily dose information.^{42,43} This approach involves linking dispensing records from the drug utilization database and converting them into records of episodes of exposure to a drug or group of drugs. The software enables the research team to determine the period of exposure in which to search for suspected adverse drug effects. Furthermore, average daily dose of drug consumed can be estimated by this methodology. These techniques have been used in a study of withdrawal seizures possibly associated with discontinuation of high-dose alprazolam.⁴² As it happened, none of the users receiving average daily alprazolam doses of 4.0 mgs or higher had a hospital admission for a seizure disorder while using the drug or during a short time after stopping use of the drug. While some high dose alprazolam users did appear to have a seizure disorder, record review showed that treatment for the seizure disorder began prior to the receipt of alprazolam.

This methodology is also important in quality assurance. For example, McFarland points out that accreditation bodies such as the National Committee for Quality Assurance are using the duration and dose of antidepressant treatment as quality assurance measures.⁴⁹ Major depressive disorder is a common condition and antidepressant drugs are often used as the main treatment modality in managed care systems.^{46,47} Information about the dosage and duration of antidepressant drug treatment derived from automated pharmacy systems can be helpful in examining quality of care.⁴⁹

THE FUTURE

The current Kaiser Permanente pharmacy databases will mature and provide a resource for evaluating long term as well as acute drug associated effects. Data from other recently implemented automated systems that capture a range of detailed clinical information can be linked with pharmacy data to conduct both retrospective and prospective evaluations of drug effects, utilization, and healthcare costs.

We anticipate that the Kaiser Permanente Medical Care Program, Northern California will continue to increase the amount of clinical and diagnostic data stored in automated systems. There are also efforts under way to make systems compatible across regions. This will facilitate collaborative pharmacoepidemiologic research of infrequently prescribed medications and/or rare side-effects or disease conditions.

In addition to aiding in the development of new diagnostic tests, advancing biomedical technologies will allow for the identification of new molecular and genetic markers that may help predict individual response to drug treatment. As promising markers are identified, we expect that investigators will incorporate them into their pharmacoepidemiology research.

Finally, the Center for Health Research at KPNW has begun the development of a "data warehouse" that will provide research-quality medical and pharmacy data drawn from several HMOs throughout the United States. Currently, the data warehouse contains limited years of data from five midsize Kaiser Permanente Divisions: Northeast (New York, Massachusetts, Connecticut, Vermont, and Rhode Island), Southeast (Georgia only), Central East (Ohio), Rocky Mountain (Colorado, Kansas, and Missouri), and Northwest (Oregon and Washington). Group Health Cooperative of Puget Sound in Seattle is also a participant. Collectively, these participating HMOs represent about 2.5 million enrollees. Efforts are under way to incorporate other Kaiser Permanente divisions into the data warehouse.

The creation and maintenance of a collaborative HMO data warehouse are subject to many challenges. The top priorities of the organizations

participating in the data warehouse are confidentiality, security, and protected use of the data. No single organization has access to the data without agreement from all participants. The data are a valuable research and business asset and are handled with extreme care.

The creation and maintenance of the data warehouse requires high level technical sophistication. Very large number of data arrive from other locations that use a myriad of operating systems, software, and data formats. Data must be "cleaned" for data quality and then integrated to allow for pooled data with minimum loss of detail. An important technical feature of the data warehouse is the "metadata." Metadata are "data about data" and includes documentation of data from all contributing sources. Metadata also include the programs used to process the integration of the data. "Data marts" are also part of the data warehouse. These are structured applications that are created to produce routine reports and other standardized output of the data.

The data warehouse is currently being used by several multisite research projects. These projects not only rely on the data warehouse as a research asset, but also helped to create its infrastructure. Among these, the Global Risk Adjustment: Payment, Estimation, Simulation study, funded by the Health Care Financing Administration, Robert Wood Johnson Foundation, and Kaiser Permanente, is the most prominent example. Researchers from Kaiser Permanente Northwest Division, Group Health Cooperative of Puget Sound, Kaiser Permanente Northeast Division, Kaiser Permanente Rocky Mountain Division, Kaiser Permanente Central East Division, and HealthPartners are collaborating on a nationwide study of Medicare payment in HMOs. The data warehouse serves as the central, research quality database.

ACKNOWLEDGMENTS

Portions of this chapter were taken from the paper by GDF, Monitoring of drug effects in outpatients: development of a program to detect carcinogenesis. pp. 55–62. In Ducrot H, Goldberg M, Hoigne R, and Middleton P (eds): Computer

Aid To Drug Therapy And To Drug Monitoring, North-Holland, Amsterdam: 1978, with permission of the publisher.

REFERENCES

1. Davis LS, Collen MF, Rubin L, *et al.* Computer-stored medical record. *Comput Biomed Res* 1968; **1**: 452–69.
2. Friedman GD, Collen MF. A method of monitoring adverse drug reactions. In: LeCam LM, Neyman J, Scott EL, eds, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. VI. Berkeley, CA: University of California Press, 1972; 367–82.
3. Friedman GD, Collen MF, Harris LE, *et al.* Experience in monitoring drug reactions in outpatients: the Kaiser-Permanente Drug Monitoring System. *J Am Med Assoc* 1971; **217**: 567–72.
4. Friedman GD, Gerard MJ, Ury HK. Clindamycin and diarrhea. *J Am Med Assoc* 1976; **236**: 2498–500.
5. Selby JV, Friedman GD, Herrinton LJ. Pharmaceuticals other than hormones. In Schottenfeld D, Fraumeni JF Jr, eds, *Cancer Epidemiology and Prevention*, 2nd edn. New York: Oxford University Press, 1996; 489–501.
6. Hoover R, Fraumeni JF Jr. Drugs. In Fraumeni JF Jr, ed., *Persons at high risk of cancer: an approach to cancer etiology and control*. New York: Academic, 1975; 185–99.
7. Friedman GD, Ury HK. Initial screening for carcinogenicity of commonly used drugs. *J Natl Cancer Inst* 1980; **65**: 723–33.
8. Friedman GD, Ury HK. Screening for possible drug carcinogenicity: second report of findings. *J Natl Cancer Inst* 1983; **71**: 1165–75.
9. Selby JV, Friedman GD. Screening prescription drugs for possible carcinogenicity: eleven to fifteen years of follow-up. *Cancer Res* 1989; **49**: 5736–47.
10. Van Den Eeden SK, Friedman GD. Prescription drug screening for subsequent carcinogenicity. *Pharmacoepidemiol Drug Safety* 1995; **4**: 275–87.
11. Friedman GD. Barbiturates and lung cancer in humans. *J Natl Cancer Inst* 1981; **67**: 291–5.
12. Friedman GD, Habel LA. Barbiturates and lung cancer: a reevaluation. *Int J Epidemiol* 1999; **28**: 375–9.
13. Olsen JH, Wallin H, Boice JD Jr, Rask K, Schulgen G, Fraumeni JF Jr. Phenobarbital, drug metabolism, and human cancer. *Cancer Epidemiol Biomarkers Prev* 1993; **2**: 449–52.
14. Habel LA, Bull SA, Friedman GD. Barbiturates, smoking, and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev* 1998; **7**: 1049–50.

15. Beard CM, Noller KL, O'Fallon WM, *et al.* Lack of evidence for cancer due to use of metronidazole. *New Engl J Med* 1979; **301**: 519–22.
16. Friedman GD. Cancer after metronidazole [letter]. *New Engl J Med* 1980; **302**: 519.
17. Goldin AG, Safa AR. Digitalis and cancer. *Lancet* 1984; **1**: 1134.
18. Friedman GD. Digitalis and breast cancer [letter]. *Lancet* 1984; **2**: 875.
19. Williams RR, Feinleib M, Connor RJ, *et al.* Case-control study of antihypertensive and diuretic use by women with malignant and benign breast lesions detected in a mammography screening program. *J Natl Cancer Inst* 1978; **61**: 327–35.
20. Friedman GD. Rauwolfia and breast cancer: no relation found in long-term users age fifty and over. *J Chron Dis* 1983; **36**: 367–70.
21. Arellano MG, Petersen GR, Petitti DB, *et al.* The California Automated Mortality Linkage System (CAMLIS). *Am J Public Health* 1984; **74**: 1324–30.
22. Ferrara A, Karter A, Gleit D, Ackerson L, Darbinian J, Selby J. Hormone replacement therapy is associated with better glucose control in a multi-ethnic population with type 2 diabetes: the Northern California Kaiser Permanente Diabetes Registry. *Diabetes* 1998; **47** (suppl 1).
23. Selby JV, Ettinger B, Swain BE, Brown JB. First 20 months experience with use of metformin for type 2 diabetes in a large health maintenance organization. *Diabetes Care* 1999; **22**: 38–44.
24. Ettinger B, Li DK, Klein R. Unexpected vaginal bleeding and associated gynecologic care in postmenopausal women using hormone replacement therapy: comparison of cyclic versus continuous combined schedules. *Fertil Steril* 1998; **69**: 865–9.
25. Levin TR, Schmittiel JA, Henning JM, Kunz K, Henke CJ, Colby CJ, Selby JV. A cost analysis of a *Helicobacter pylori* eradication strategy in a large health maintenance organization. *Am J Gastroenterol* 1998; **93**: 743–7.
26. Quesenberry CP Jr, Caan B, Jacobson A. Obesity, health services use, and health care costs among members of a health maintenance organization. *Arch Intern Med* 1998; **158**: 466–72.
27. Ettinger B, Pressman A, Bradley C. Comparison of continuation of postmenopausal hormone replacement therapy: transdermal versus oral estrogen. *Menopause* 1998; **5**: 152–6.
28. Lieu TA, Quesenberry CP, Sorel ME, Mendoza GR, Leong AB. Computer-based models to identify high-risk children with asthma. *Am J Respir Crit Care Med* 1998; **157**: 1173–80.
29. Friedman GD. Phenylbutazone, musculoskeletal disease, and leukemia. *J Chron Dis* 1982; **35**: 233–43.
30. Hiatt RA, Bawol RD, Friedman GD, *et al.* Exogenous estrogens and breast cancer after bilateral oophorectomy. *Cancer* 1984; **54**: 139–44.
31. Friedman GD. Multiple myeloma: relation to propoxyphene and other drugs, radiation, and occupation. *Int J Epidemiol* 1986; **15**: 424–6.
32. Goldhaber MK, Selby JV, Hiatt RA, Quesenberry CP Jr. Exposure to barbiturates in utero and during childhood and risk of intracranial and spinal cord tumors. *Cancer Res* 1990; **50**: 4600–3.
33. Hiatt RA, Tolan K, Quesenberry CP Jr. Renal cell carcinoma and thiazide use: a historical case-control study. *Cancer Causes Control* 1994; **5**: 319–25.
34. Petitti DB, Sidney S, Bernstein A, Wolf S, Quesenberry CP, Ziel HK. Stroke in users of low dose oral contraceptives. *New Engl J Med* 1996; **335**: 8–15.
35. Sidney S, Petitti DB, Quesenberry CP, Klatsky AL, Ziel HK, Wolf S. Myocardial infarction in users of low-dose oral contraceptives. *Obstet Gynecol* 1996; **88**: 939–44.
36. 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.
37. Chen RT, Glasser JW, Rhodes PH, Davis RL, *et al.* Vaccine Safety Datalink project: a new tool for improving vaccine safety monitoring in the United States. *Pediatrics* 1997; **99**: 765–73.
38. Corelle C, Bennett M, Johnson RE, McFarland B. Using SAS software in pharmacoepidemiologic research: identifying episodes of drug use and determining average daily dose. *SUGI 18: Proceedings of the 18th Annual SAS Users Group International Conference, New York, 1993*. Canj, NC: SAS; 549–51.
39. Davis RL, Marcuse E, Black S, *et al.* MMR2 immunization at 4 to 5 years and 10 to 12 years of age: a comparison of adverse clinical events after immunization in the Vaccine Safety Datalink project. *Pediatrics* 1997; **100**: 767–71.
40. Johnson RE, Mulooley JP, Valanis BG, McFarland BH, Andrews EB, Tilson HH. Method of examining oral acyclovir use for adverse events. *J Clin Res Pharmacoevidiol* 1991; **5**: 331–46.
41. Johnson RE, McFarland BH. Antipsychotic drug exposure in a Health Maintenance Organization. *Med Care* 1993; **31**: 432–44.
42. Johnson RE, McFarland BH, Corelle CA, Woodson GT. Estimating daily dose for pharmacoepidemiologic studies: alprazolam as an example. *Pharmacoevidiol Drug Safety* 1994; **3**: 139–45.
43. Johnson RE, McFarland BH. Lithium use and discontinuation in an HMO. *Am J Psychiatry* 1996; **153**: 993–00.
44. Johnson RE, McFarland BH, Woodson G. Whither. Triazolam. *Med Care* 1997; **35**: 303–10.
45. Johnson RE, McFarland BH, Nichols G. Changing patterns of antidepressant use in an HMO. *Pharmacoeconomics* 1997; **11**: 274–86.

46. McFarland BH. Cost-effectiveness considerations for managed care systems: treating depression in primary care. *Am J Med* 1994; **97**: 47–58.
47. McFarland BH. New drug class for comorbid depression. *Am J Managed Care* 1996; **2**: 721–5.
48. McFarland BH. Comparing period prevalences. *J Clin Epidemiol* 1996; **49**: 473–82.
49. McFarland BH. Automated pharmacy databases and behavioral health care quality assurance. *Eval Rev* 1997; **21**: 371–8.
50. Weiss SR, McFarland BH, Corelle C, McCloskey C, Vega A. Patterns of drug use in pregnancy. *Pharmacoepidemiol Drug Safety* 1997; **6**: S69.
51. Weiss SR, McFarland BH, Burkhart GA, Ho PTC. Cancer recurrences and second primary cancers after use of antihistamines or antidepressants. *Clin Pharmacol Ther* 1998; **63**: 594–9.

17

Harvard Pilgrim Health Care/Harvard Vanguard Medical Associates

K. ARNOLD CHAN

Department of Epidemiology, Harvard School of Public Health, Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

RICHARD PLATT

Harvard Pilgrim Health Care, Department of Ambulatory Care and Prevention, Harvard Medical School, Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

INTRODUCTION

Harvard Pilgrim Health Care (HPHC), a 1.1 million member nonprofit health maintenance organization (HMO) in New England, has a long history of public domain research that uses automated record linkage methods in a wide range of epidemiologic and health services research (including pharmacoepidemiology, drug policy, and other health services research) and in support of clinical trials. Approximately 90% of HPHC members have prescription drug benefits that offers them a strong incentive to fill their prescriptions at a pharmacy that reports the dispensing to the HMO. These computerized pharmacy records are fully linkable to automated ambulatory care, emergency room visit, and hospital admission records. In addition, a fully automated ambulatory medical records system is in place for approximately 300 000 members. This system captures all encounters between members and providers, including telephone calls; diagnosis, test, and treatment codes for each

ambulatory encounter; and providers' full-text notes and dictations. It is possible to search these automated office records for coded clinical conditions, tests, and prescriptions and to link these data with emergency room and hospital discharge diagnoses. This chapter reviews the use of HPHC data for pharmacoepidemiology research.

DESCRIPTION

OVERVIEW OF HPHC/HVMA

HPHC is the largest HMO in New England, with approximately 1.1 million members as of January 2000. This mixed-model, nonprofit HMO, which was formed in 1994 from the merger of the Harvard Community Health Plan and Pilgrim Health Care, provides primary care and all major medical subspecialty services. In 1998, the 300 000-member former staff-model component became a multi-specialty group, Harvard Vanguard Medical

Associates (HVMA). Most HVMA patients are HPHC members.

Pharmacy dispensing information is available from 1988 (when the health plan population was approximately 250 000) through the present; this information includes data on 90% of HPHC members with pharmacy benefits. In addition, prescribing information is available for 1969 onward for a cumulative total of roughly one million individuals (including 300 000 current members) with automated ambulatory medical records. The prescription drug benefit provides prescription drugs for a copayment (usually less than \$10 for a month's supply) that provides a strong incentive to use a pharmacy that provides automated dispensing data to HPHC. For each dispensing the following data are recorded: a unique patient identifier, gender, date of birth, dispensing date, prescriber identifier, generic drug name, dosage strength, amount dispensed, route, American Hospital Formulary System (AHFS) code, National Drug Code (NDC), member's contract number, and dependency code (noting the relationship between family members and contract holders).

DEMOGRAPHICS OF HPHC MEMBERS

The median age of HPHC members in 1998 was 31 years. Slightly more than half (53%) of members were female. Twenty-seven% of members were age 17 or younger, and 6.7% were 65 or older. Among HPHC members who identified their race during a survey, 73% were White, 16% were African-American, 2% were Asian, and 9% were Hispanic/other. HPHC has active Medicare and Medicaid programs.

HPHC AUTOMATED RECORDS

Membership Records

Membership files include demographic and health plan coverage information for each HPHC member. Date of birth, gender, dates of membership (including all starts, stops, and changes in coverage or benefit status during each membership interval), and most recent residential zip code are reliably recorded. Race, marital status, education, and occupation are less reliably recorded. This record

also includes family membership information that allows linkage of family members and data on the organization through which membership was obtained (private corporation, state employer, Medicare, Medicaid, or individual).

Ambulatory Encounter Records

All ambulatory encounters generate claims with up to two primary and secondary diagnosis codes and one procedure code as well as information about the provider and dates of encounter. Diagnoses are coded according to the *International Classification of Disease*, 9th edition (ICD-9), and procedures are coded according to the Current Procedural Terminology (CPT).

In addition, there are fully automated medical records for HVMA patients. For these individuals, patients' vital signs and an unlimited number of diagnoses per encounter are captured during encounters (including telephone calls). Providers' notes and dictations are transcribed in their entirety into the automated medical record. Results of most laboratory tests, diagnostic imaging procedures, and other tests performed within and outside HVMA facilities are entered directly into the automated record and are linked with the patient encounter during which they were ordered. Hard copy medical records are maintained for information that cannot readily be incorporated into the automated medical record, such as electrocardiograms, signed informed consent forms, and medical records from prior health care providers.

Records for Emergency Room Visits, Hospitalizations, and Other Medical Services

All emergency room visits and hospitalizations generate claims (form UB82 was used through 1993 and form UB92 thereafter) with the hospital name, dates of hospitalization, discharge diagnoses, and procedures performed. The largest single inpatient facility for HPHC/HVMA patients is Brigham and Women's Hospital, where detailed automated inpatient records are available.¹ Other medical services, including long-term care and home care provided by physicians, nurses, or other private vendors, are reported to HPHC on UB92 or HCFA 1500 forms (which contain up to four ICD-9

coded diagnoses and six coded procedures) and are also available for automated searches.

Linkage to Data Sources Outside HPHC

The research staff can link information about the HMO's members to external data sources, such as state vital statistics registries and the records of the state board of registry in medicine.

USE OF HPHC DATA FOR PHARMACOEPIDEMIOLOGY RESEARCH

In many pharmacoepidemiology studies, use of a certain drug or a drug class is the "exposure" information and clinical events after the initial exposure are the "outcome" information. Drug prescribing or dispensing is sometimes the outcome of interest in studies of the effect of drug policies. Potential confounders and effect modifiers may include demographic factors, concomitant medications, and concurrent diseases. All these data are routinely collected for all members.

Use of Pharmacy Data

Most pharmacoepidemiology studies requiring complete drug-exposure information are restricted to individuals with prescription drug coverage. Members who have filled a prescription for a given drug can be identified by the brand name, generic name, AHFS class, NDC, or any combination of these identifiers. Further linkage with demographic records can restrict the exposed population to those who have continuous benefit coverage for prescription drugs during a certain period. Concomitant medications dispensed to the same individuals would provide information on concurrent diseases and the severity of these diseases.

Use of Diagnosis and Procedure Data

An outcome after drug exposure can be identified by information from several sources, including ambulatory diagnoses, hospital discharge diagnoses, radiology examinations, laboratory examinations, and procedures such as endoscopic examinations. Outcome identification usually in-

volves a screening step in which one or more diagnosis and/or procedure codes are sought in automated ambulatory or hospital records and selected full-text records are reviewed to confirm the diagnosis. Information on medical history or concurrent illness can be obtained during the manual record review.

Coded Diagnosis Information

The ICD-9 diagnoses associated with ambulatory encounters or hospitalizations for all members can be identified from automated claims files. In order to use coded information for research purposes, several types of fixed length record are usually created for patients identified by such searches. There is a single record containing demographic data for each individual, a fixed length record for each ambulatory encounter, a separate clinical encounter record that contains prenatal information, and administrative records about each service that a member obtains, including care provided within physicians' offices, at hospitals and emergency rooms, and by other providers, such as home health care agencies. All records include a unique patient identifier. If there are more diagnoses, tests, or treatments than can be recorded in a single fixed length encounter record, continuation records are created to capture all coded information. Records of laboratory testing contain the test type, test date, and test result. For members with automated ambulatory records, a fully automated record database can be searched in cases with diagnosis codes or characteristics of interest.

Review of Full-Text Medical Records

Selected parts of the full-text medical records are ordinarily reviewed to confirm coded information or to obtain details on the outcomes of interest and/or on potential confounders and effect modifiers. Record reviews have been performed routinely for research and quality assurance purposes. Hospital discharge summaries and full-text medical records for all members are obtained from participating hospitals and manually reviewed. Full-text ambulatory records are also available from medical care providers. For HVMA members with automated

ambulatory records, full-text records are available in electronic format, and text filters can be used to highlight specific text strings, thereby allowing efficient manual review. Typically, 75% of hospital records and nonautomated ambulatory records are accessible for review for research purposes; all automated ambulatory records are usually available.

Figure 17.1 shows an example of a full-text record retrieved during a study of asthma medications. The full-text records of these encounters were printed, with text strings indicating the conditions of interest highlighted. Words of interest are highlighted with automated text processing methods, i.e., by placing three asterisks before them and underlining the line in which they occur.

<p>[HPHC member number] ENCOUNTER--[MM/DD/YY]</p> <p>OBJECTIVE DATA TEMP 98.4</p> <p>PROBLEMS <u>G992 ***COUGH</u> X10 DAYS W/S.T. BOTH RESOLVING.THROAT SLIGHTLY INJECTED.2 SMALL SOFT POSTERIOR NODES PALPABLE.CHEST CLEAR, AFEB, <u>***COUGH MAINLY DRY</u></p> <p>THERAPIES M992 SORE THROAT LOZENGES R001 SYMPTOMATIC RX FLDS,REST</p> <p>SITE: [XXXXXXXXX] TYPE: SCHEDULED</p> <p>[Name of care-giver], RN/NP/SMS</p>
<p>ENCOUNTER--[MM/DD/YY]</p> <p>PROBLEMS A128 PRESCRIPTION REFILL ONLY <u>TAVIST D. & VENTOLIN ***INHALER PER DR. [XXXXXX]</u></p> <p>THERAPIES A126 CLEMASTINE (TAVIST) TAVIST D #60 1 PO BID NRF <u>J134 ALBUTEROL (PROVENTIL, VENTOLIN) VENTOLIN #1 ***INHALER PRN AS DIRECTED</u> NRF</p> <p>SITE: [XXXXXXXXX] URGENT CARE-EVENING (AFTER HOURS) TYPE: TELEPHONE</p> <p>[Name of care-giver], RN/VRW</p>

Figure 17.1. Excerpts of automated ambulatory medical record from two encounters for an HVMA member, after processing by a text filter designed to facilitate review for asthma by focusing on symptoms (cough) and prescription (a refill over the telephone).

The highlighting shown in Figure 17.1 can be applied to both coded information and free text. Similar methods are used to mask the names of drugs being studied. These selective text printing capabilities also can be used to enhance the efficiency of automated reviews, increase confidentiality by obscuring most occurrences of the patient's name, and reduce the opportunity for reviewer bias by masking the names of specific drugs.

STRENGTHS

The HPHC/HVMA research environment offers several advantages for pharmacoepidemiology studies.

- The availability of data from 1969 onward allows certain longitudinal studies.
- Since most members have prescription drug benefits, exposure information is likely to be complete for most such drugs.
- Full-text ambulatory and hospital records can be used for research. Automated ambulatory records, which are available for part of the membership, are an extremely rich research resource, providing essentially complete, highly efficient automated access to a level of clinical detail that is often unavailable or available only with great effort.
- Laboratory test results can be obtained from automated sources for members with automated ambulatory records.
- A full-time research staff both has expertise in performing record linkage research and also has access to primary data sources.

WEAKNESSES

- The principal limitation of the data source, as of most HMO based databases, is the substantial turnover of the membership. The annual disenrollment rate is approximately 14% among all members. However, the attrition rate is much lower for members who have been enrolled for more than three years and for those with chronic conditions, such as asthma

or hypertension. It is difficult to ascertain outcomes after membership ceases.

- Data required for research reside in diverse computer systems. Accessing these systems and obtaining information in usable form require research staffs with sophisticated programming skills and experience with these datasets.

PARTICULAR APPLICATIONS

OVERVIEW OF RESEARCH STRATEGY

The pharmacy dispensing and claims data support a variety of study designs, including cohort and nested case-control studies (Table 17.1). For cohort studies, exposure can be based on drug exposure or clinical status. Drug exposure is usually defined on the basis of the dispensing of a drug of interest. It is possible to select drug exposed cohorts with desired age, sex, and membership characteristics and to identify individually matched or group matched unexposed individuals for comparison. Usually the population is restricted to individuals who have prescription drug benefits and are therefore most likely to have filled prescriptions at HVMA or contract pharmacies.

EXAMPLES OF PHARMACOEPIDEMIOLOGY STUDIES CONDUCTED AT HPHC

Use of Pharmacy Dispensing Records and Hospital Discharge Diagnoses To Evaluate Potential Adverse Drug Reactions

A retrospective cohort study of the association between alendronate and gastroduodenal perforation, ulcer, or bleeding was conducted at HPHC and seven other HMOs. Health plan members to whom alendronate was dispensed during a specific interval were identified from drug dispensing records. Subjects were also required to have pharmacy benefits and to have been members for at least one year before the start of alendronate exposure. A comparison group who did not have

Table 17.1. Examples of pharmacoepidemiology projects performed using HPHC data

-
- Serum sickness associated with oral antibiotics
 - *C. difficile* associated diarrhea following outpatient antibiotic treatment
 - β -adrenergic receptor blocker-associated depression
 - Sudden death, syncope and arrhythmias following prescription antihistamines
 - Varicella zoster epidemiology, complications, and patterns of care
 - Ibuprofen and skin and soft tissue superinfections in children with varicella
 - H₂ blocker associated hemolytic anemia
 - Cutaneous drug reactions and other dermatologic conditions in HIV infection
 - Isotretinoin prescribing practices
 - Use and effectiveness of asthma therapies
 - Use and effectiveness of antihypertensives
 - Use of drugs with high per patient costs
 - Use of lipid lowering agents in primary care settings and reasons for discontinuation of use
 - Incidence of severe skin eruptions after lamotrigine use
 - Identification and epidemiology of surgical site infections after hospital discharge
 - Alendronate and gastroduodenal perforation, ulcer and bleeding
-

alendronate dispensed but who were comparable in terms of age and sex distribution, pharmacy benefit, and duration of membership was randomly selected. Investigators then searched the records of exposed and unexposed subjects for hospital discharge diagnosis codes that may represent upper gastrointestinal perforation, ulcer, or bleeding. Hospital records were reviewed to confirm the clinical event of interest. Incidence rates for exposed and unexposed individuals were computed from the confirmed events and eligible person time. For alendronate exposed individuals, separate incidence rates were computed for all time after alendronate dispensing and time during which the individuals had an adequate supply of drug. Adjusted incidence rate ratios were derived, with covariates obtained from automated data sources, such as age, sex, and medical care (including exposure to a wide array of prescription drugs that are markers of specific chronic diseases) received during the year before observation for the

outcomes of interest began.² A complementary assessment of confounding by indication, which would exist if osteoporosis were independently associated with gastroduodenal perforation, ulcer, or bleeding, was performed by assessing the rate of these events among age matched individuals with nonpathologic fractures, which were chosen as the best available marker of osteoporosis.

Use of Automated Ambulatory, Pharmacy Dispensing, and Hospitalization Records To Evaluate Drug Effectiveness

A retrospective cohort study among asthmatics was conducted in the group of HPHC members who had automated ambulatory records for 1991 through 1994 to evaluate risk factors associated with hospitalizations for asthma. Asthma was defined as at least one asthma diagnosis during an ambulatory encounter, a hospitalization, or an emergency room visit. The study population consisted of 16 941 HPHC members. After adjusting for age, sex, other asthma medications, and amount and type of ambulatory care for asthma, use of inhaled steroids was found to be associated with a 50% decrease in the risk of hospitalization due to asthma attacks.³ The large sample size and efficient data collection in this study illustrate the strength of research based on automated medical records.

Use of Automated Dispensing, Ambulatory, and Hospital Data To Evaluate Adherence to Drug Therapy

A study of lipid lowering agents at HPHC and another HMO utilized drug dispensing data to identify a cohort of users of these drugs, to evaluate the duration of drug use, and to characterize the reasons for drug discontinuation. Followup with automated data showed that the effectiveness and tolerability of these agents were poorer in general use than had been suggested by clinical trials.⁴

Another study evaluated adherence to antihypertensive therapy. Treated hypertensive patients were identified among HPHC members with automated ambulatory medical records. Eligible subjects were

identified on the basis of measured blood pressure exceeding specified thresholds, a diagnosis of hypertension, and dispensing of antihypertensive therapy for at least six months. All patient selection was based on automated information. Subjects answered questionnaires and underwent three months of electronic adherence monitoring and pill counts. Automated ambulatory medical records and pharmacy dispensing records were reviewed to assess refilling frequency and blood pressures. Correlations between different methods for the assessment of adherence were estimated.⁵ Since the entire population of eligible people from whom the sample was recruited was known, it was possible to estimate the generalizability of the information obtained.

Use of Coded Ambulatory Diagnoses, Full-Text Medical Records, and Electrocardiograms Together with Pharmacy Records

Events considered to be either potential precursors or consequences of *torsades de pointes*, a form of ventricular tachycardia, that follows exposure to terfenadine or other antihistamines, were studied.⁶ These outcomes included sudden death, syncope, ventricular arrhythmia, and QTc interval prolongation. Exposures were defined as dispensing of an antihistamine of interest. Most syncope episodes were diagnosed by a search of the automated medical record for diagnoses such as syncope and seizure and for diagnostic tests including electrocardiograms and Holter monitoring. Hospital records were also searched. Deaths were identified through state vital statistics records. Arrhythmias and QTc interval prolongation were determined by independent review of all electrocardiograms obtained during exposure by a cardiologist. Novel features of this study included the ability to identify many episodes of syncope, even those reported by telephone, and then to review the providers' notes to confirm the diagnosis. In addition, it was possible to obtain approximately 90% of electrocardiograms. Sixty first episodes of syncope or arrhythmia were identified during 1.9 million antihistamine exposure days (0.86 million for

terfenadine, 1.04 million for other agents). Comparison of terfenadine with other antihistamines revealed a risk ratio of 0.86, with a 95% confidence interval of 0.52–1.44. Furthermore, there was no difference in prolongation of the rate adjusted QT interval.

Use of Abnormal Laboratory Test Results To Study a Rare Disease with Common Presenting Symptoms

The intent of this study was to determine the burden of *C. difficile* associated diarrhea and its association with specific antibiotics. Since positive assays are uncommon relative to diarrhea diagnoses, the identification process began with a search for all positive *C. difficile* toxin assays performed in outpatient and hospital laboratories among HPHC members with automated ambulatory records.⁷ Ambulatory and hospital records were reviewed to determine which positive laboratory tests were associated with diarrhea and colitis and which cases were hospital acquired. Since the number of person years at risk was known, it was possible to estimate the incidence rate of community acquired symptomatic disease. In addition, antibiotic-specific attack rates were determined by identifying all prescriptions dispensed from HMO pharmacies to individuals with pharmacy benefits, assigning risk periods associated with these exposures, and determining which of these risk periods were associated with *C. difficile* associated diarrhea. The overall incidence rates were 7.7 cases per 100 000 person-year and 6.7 cases per 100 000 antibiotic risk periods. More than 80% of cases were identified only in ambulatory records. For most antibiotics, the risk was between 1 in 1000 courses and 1 in 100 000 courses. Disease was significantly more common among adults than among children. A few antibiotics were associated with a significantly higher age adjusted risk than was seen in a reference group of ampicillin/amoxicillin recipients, although the limited number of cases associated with each antibiotic precluded precise estimation of the magnitude of the excess risk.

Use of Prescribing Records To Study Physicians' Behavior

A study focusing on prescribing practices used prescribing information from clinicians' records rather than pharmacy dispensing records.⁸ The goal was to assess how frequently physicians determined women's pregnancy status before prescribing isotretinoin, a teratogen. The automated medical record was used to identify isotretinoin prescriptions among women between the ages of 15 and 44 years. Women in this group who had undergone tubal ligation or hysterectomy were excluded. Automated medical records were used to identify pregnancy tests performed within 14 days before the date of prescription. Results were analyzed for the periods before and after FDA advisory committee hearings on isotretinoin and after the manufacturer's implementation of a broad based education program for physicians. Overall, prescribing declined by 45% and pretreatment pregnancy testing increased from 51% of courses to 58% of courses. There were two different patterns of change: physicians who initially tested frequently reduced their prescribing by 94%, while physicians who initially tested infrequently increased their testing frequency from 6 to 56% but also slightly increased their prescribing frequency. These findings suggest that educational interventions may have disparate effects on prescribers that can be identified only by prescriber level analyses.

Use of Coded Diagnoses and Test Results To Identify a Cohort Requiring Intensive Review

Both coded diagnoses and test results were used to identify a cohort of potentially HIV infected individuals.⁹ Full-text ambulatory and hospital records were reviewed to confirm the diagnosis and to identify cutaneous diseases, including those attributed to drugs. The drug exposures of these individuals were identified from the automated pharmacy records, and attack rates were calculated. The risk of cutaneous diseases was substantially higher among HIV infected individuals than among HIV uninfected individuals. The risk increased with the clinical stage of HIV infection.

Other Studies

Additional published studies from HPHC include an assessment of the relative risk of serum sickness in children following exposure to commonly used antibiotics;¹⁰ diagnosis of depression after prescription of propranolol;¹¹ occurrence of hemolytic anemia following exposure to selected H₂ antagonists;¹² surveillance for surgical site infection using a combination of drug exposures, diagnoses, and treatments to identify individuals who may have been infected;¹³ and ibuprofen and skin and soft tissue superinfections in children with varicella.¹⁴

APPROPRIATE ROLE OF HPHC/HVMA DATA RESOURCES IN PHARMACOEPIDEMOLOGY

This record linkage system is useful for assessing clinical and laboratory outcomes after drug exposures. In addition to automated drug dispensing records and claims system records containing ICD-9 diagnosis and procedure codes that allow studies comparable to those of other automated record linkage systems, the HVMA system includes automated full-text ambulatory medical records, automated data on ambulatory blood pressures and other vital signs, coded physicians' prescribing data, and automated ambulatory laboratory test results. The availability of these data permits several types of investigation, including studies of prescribing practices, assessment of prescription filling behavior, and evaluation of conditions that are identified and managed principally in the ambulatory setting. The ability to link outpatient and inpatient drug exposures for some members also allows researchers to assess post-discharge outcomes of inpatient drug exposures and to assess more completely drug exposures either in or out of the hospital.

THE FUTURE

In the future, we anticipate developing more studies that combine "classic" pharmacoepidemiology methods with a broad array of drug policy and other health services research ap-

proaches. Pharmacoepidemiology methods will play a larger role in HMO populations both in observational studies and in clinical trials of disease management guidelines and therapeutic interventions. We expect to place more emphasis on obtaining selected information (e.g., symptom data, quality-of-life assessments, or information on health beliefs) or specimens (e.g., blood samples for genotyping) directly from members as part of specific studies. We also expect to extend our collaborations with other research groups that will allow investigations encompassing the substantially larger populations (5 million–15 million individuals) required for the timely investigation of many topics of scientific and public health interest.

REFERENCES

1. Platt R, Stryker WS, Komaroff AL. Pharmacoepidemiology in hospitals using automated data systems. *Am J Prev Med* 1988; **4** (suppl 2): 39–47.
2. Donahue JG, Andrade S, Beck A, Boles M, Buist D, Chandler J, Chase G, Ettinger B, Goodman M, Guess, H, Gurwitz J, Lacroix A, Levin TR, Platt R. Multi-HMO assessment of alendronate and gastroduodenal perforation, ulcer or bleeding. *Drug Inf J* 1999; in press.
3. Donahue JG, Weiss ST, Goetsch MA, Livingston JM, Griener D, Platt R. Inhaled steroids and hospitalization for asthma. *J Am Med Assoc* 1997; **277**: 887–91.
4. Andrade SE, Walker AM, Gottlieb LK, Hollenberg NK, Testa MA, Saperia GM, Platt R. Discontinuation of antihyperlipidemic drugs: do rates reported in clinical trials reflect rates in primary care settings? *New Engl J Med* 1995; **332**: 1125–31.
5. Choo PW, Rand CS, Inui TS, Lee M-LT, Cain EM, Cordeiro-Berault M, Canning C, Platt R. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. *Med Care* 1999; **37**: 846–57.
6. Hanrahan JP, Choo PW, Carlson W, Greineder D, Faich GA, Platt R. Antihistamine associated ventricular arrhythmias and QTc interval prolongation. *Ann Epidemiol* 1995; **5**: 201–9.
7. Hirschhorn LR, Trnka Y, Onderdonk A, Platt R. Epidemiology of community-acquired *Clostridium difficile* associated diarrhea. *J Infect Dis* 1994; **169**: 127–33.
8. Heckbert SR, Mtshali G, Platt R. Effect of physician and patient education efforts on prescribing of isotretinoin to women of childbearing age. *Pharmacoepidemiol Drug Safety* 1993; **2**: 217–21.
9. Coopman SA, Johnson RA, Platt R, Stern RS. Cutaneous disease and drug reactions in HIV infection: a population-based study. *New Engl J Med* 1993; **328**:1670–4.
10. Heckbert SR, Stryker WS, Coltin KL, Manson JE, Platt R. Serum sickness following exposure to cefaclor or amoxicillin in an HMO population. *Am J Epidemiol* 1990; **132**: 336–42.
11. Gerstman B, Jolson H, Bauer M, Choo P, Livingston J, Platt R. Occurrence of depression in new users of beta-blockers and selected antihypertensives. *J Clin Epidemiol* 1996; **49**: 809–15.
12. Choo PW, Goldberg JH, Platt R. Ranitidine associated autoimmune hemolytic anemia in a health maintenance organization population. *J Clin Epidemiol* 1994; **10**: 1175–9.
13. Sands K, Vineyard G, Platt R. Surgical site infections occurring after hospital discharge: epidemiology and methods for detection. *J Infect Dis* 1996; **173**: 963–70.
14. Choo PW, Donahue JG, Platt R. Ibuprofen and skin and soft tissue superinfections in children with varicella. *Ann Epidemiol* 1997; **7**: 440–5.

18

UnitedHealth Group

DEBORAH SHATIN AND CAROL DRINKARD

Center for Health Care Policy and Evaluation, UnitedHealth Group, Minneapolis, MN, USA

ANDY STERGACHIS

Department of Pharmacy and Department of Epidemiology, University of Washington, Drugstore.com, Bellevue, WA, USA

INTRODUCTION

UnitedHealth Group is a diversified company providing health and well-being services to more than 13 million members throughout the United States. UnitedHealth Group established the Center for Health Care Policy and Evaluation (the Center) in 1989 as a private sector research institute with an independent research agenda. The Center was created to foster objective research for UnitedHealth Group affiliated health plans and specialty companies as well as federal, state, and private sector clients.

UnitedHealth Group provides a link to over 40 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. This access to medical management information data reflecting a broad cross section of the population gives researchers at the Center and their collaborators unique research opportunities. The Center uses a subset of the electronic claims data, the research databases (RDB), to

conduct a variety of customized studies including pharmacoepidemiology, quality and performance, health outcomes, and cost-effectiveness, often in collaboration with sponsors, academicians, and/or government agencies. For example, prescription drugs, biologics, and medical devices can be monitored for adverse events using postmarketing surveillance methods. This chapter describes UnitedHealth Group and the research databases that link various health care service data files, advantages and disadvantages of utilizing these databases for pharmacoepidemiology studies, examples of such studies, and future directions of post-marketing surveillance and pharmacoepidemiology studies in this setting.

DESCRIPTION

OVERVIEW OF UNITEDHEALTH GROUP

UnitedHealth Group is a health care management company serving purchasers, consumers, managers, and providers of health care. Founded in

1974, UnitedHealth Group serves more than 13 million persons through a continuum of health care and specialty services. These services include health maintenance organizations (HMOs), point of service (POS) arrangements, preferred provider organizations (PPO), and managed indemnity programs, Medicaid and Medicare managed care programs, and the American Association of Retired Persons (AARP) insurance programs. Other services include managed mental health and substance abuse services, utilization management, specialized provider networks, third-party administration (TPA) services, employee assistance services, managed pharmacy services, and information systems.

UnitedHealth Group has approximately 290 000 participating providers, 3050 participating hospitals, and 53 000 participating retail pharmacies via a pharmacy benefits management company. Membership varies over time and includes approximately 5 800 000 commercial health plan members as well as 500 000 Medicaid and 450 000 Medicare beneficiaries. Unique member identifiers allow for tracking across enrollment periods so a member can be followed through disenrollment and re-enrollment.

Although managed care is often viewed as a unitary concept and was the initial HMO model,¹ plan structures vary and range from staff or group models to independent practice associations (IPAs). UnitedHealth Group affiliated health plans are typically IPA models, with open access to a wide network of providers. Certain UnitedHealth Group affiliated health plans have offered gatekeeper or capitated models, although the emphasis remains primarily open access and discounted fee-for-service models.

THE CENTER FOR HEALTH CARE POLICY AND EVALUATION

The Center for Health Care Policy and Evaluation was recommissioned in 1997 by UnitedHealth Group with an independent research agenda to generate objective information that will contribute to improving health care delivery, advancing the health of the public, and better informing public policy makers, health care organizations, and

government agencies. Initially, the Center collaborated with a consortium of pharmaceutical companies through the Medical Therapeutics and Outcomes Project (MTOP) to compare accuracy of administrative claims data and medical record information.² Other early Center research resulted in the development of a Quality Screening Management (QSMTM) system as the first quality measurement system implemented across health plans.³ Elements of QSMTM were incorporated into The Health Plan Employer Data and Information Set (HEDIS), which compares quality across health plans. Further, the Center developed health plan "Report Cards" to evaluate plan performance including quality of care, cost reduction, operating efficiency/administrative costs, and consumer satisfaction.⁴ The Center now collaborates with government agencies, universities, other health plans, and research organizations. An overview of health services research and evaluation activities at the Center is described in more detail elsewhere.⁵

Most recently, the Center has collaborated with the US Food and Drug Administration (FDA) as a Cooperative Agreement site with consultation from the Department of Pharmacy at the University of Washington in Seattle. Since 1994, the Center has used the large population based research databases (RDB) to evaluate drug usage and study adverse events that are first identified through the FDA Spontaneous Reporting System (SRS).

RESEARCH DATABASES (RDB)

The research databases are comprised of current and historical medical and pharmacy administrative claims data submitted by 12 UnitedHealth Group affiliated health plans that are geographically diverse, including health plans in the Northeastern, Southeastern, Midwestern, and Western regions of the US. The RDB administrative claims databases are large longitudinal databases, consisting of more than nine years of data from 1990 to the present. In 1997, there were approximately 3.5 million members and 2.6 million member years, representing commercial, Medicaid, and Medicare populations. For the purpose of postmarketing

drug surveillance, analyses are typically restricted to those members having a drug benefit, approximately 93% of commercial members and most Medicaid members in the RDB. Since Medicare drug benefits vary depending on the plan, pharmacy files may not capture all prescribed drugs if beneficiaries reach the drug benefit limit.

The research databases are used to link files longitudinally and are organized from the following components (Table 18.1):

- *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.
- *Medical claims.* A claim form must be submitted by a health care provider in order 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 service, including specialty, preventive, and office-based treatment. Health plan providers submit claims either by mail or electronically.
- *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 pharmacy services covered by the health plan and filled by the member. In addition to a patient identification number, each pharmacy claim specifies the pharmacy, drug name, date dispensed, dosage of medication dispensed, duration of the prescription in days, and quantity dispensed.
- *Provider data.* A provider file contains data on the health plan’s participating physicians and other providers, including provider type and location, as well as physician specialty or

Table 18.1. Summary description of data elements included in the UnitedHealth Group research databases

Type	Selected administrative data elements
Membership Data	HMO identifier Member identifier Date of birth Gender Date of enrollment Date of disenrollment
Medical claims data: ambulatory	HMO identifier Member identifier Provider specialty Provider identifier Date service provided ICD-9 diagnosis code CPT-4 procedure code Place of service Amount claimed
Medical claims data: institutional	HMO identifier Member identifier Provider identifier Beginning date of service Ending date of service Principal ICD-9 diagnosis code Other ICD-9 diagnosis codes CPT-4 procedure codes Revenue code Amount claimed
Pharmacy claims data	HMO identifier Member identifier Prescribing physician identifier Pharmacy identifier NDC code Generic code Drug name Drug strength Dosage form Quantity of drug dispensed Days supply Date filled Amount claimed
Provider data	HMO identifier Provider specialty Provider identifier Prescribing physician identifier Provider name Provider mailing address

subspecialty. A unique provider identification number is assigned to each provider and institution. Precautions are taken to protect the identity of providers.

The components of these databases are directly applicable to addressing pharmacoepidemiology questions regarding drug exposures and adverse drug events. The various files described above are incorporated into software designed by UnitedHealth Group to facilitate the investigation of pharmacoepidemiology research questions. Research capabilities include:

- *Performing record and file linkages.* Enrollment, medical claims, pharmacy claims, and physician records can be integrated by linking members' discrete records, inpatient and outpatient claims to pharmacy claims, and all claims to member and provider data. These linkages allow for the analyses of episodes of care and the investigation of procedures and treatments regardless of location rendered.
- *Constructing longitudinal histories.* Information on diagnosis, treatments, and the occurrence of adverse clinical events, as coded on claims, can be tracked across time. To facilitate this process, the Center constructs more complete longitudinal histories by tracking members who have had multiple enrollment periods and identification numbers with a plan. Similarly, programs have been written to combine data from providers with multiple identification numbers.
- *Identifying denominators to calculate rates.* UnitedHealth Group's administrative 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 file, all individuals eligible to receive medical services or outpatient pharmacy services are identified. These populations can be defined by age, sex, or benefit status, period and duration of enrollment, or geography. Through the medical and pharmacy claims, subpopulations can be identified for calculating the prevalence and incidence of specific diseases or utilization of particular treatments.
- *Identifying treatment at a particular point in time.* The ability to identify and track treatment is a critical function in pharmacoepidemiology research. For instance, specific treatments at a particular point in time can be identified using 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 provider. For example, the researcher may want to locate a medical record for collection of detailed information not captured in the claims data. Because members receive care from multiple providers, logic has 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 following data elements necessary to calculate person time at risk: (i) dates prescriptions were filled, amount dispensed, and days supply, and (ii) the period and duration of enrollment for each member. The drug strength, amount dispensed, and days supply fields can be used to determine 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 members have been enrolled in the plan. Disenrollment from the health plan is a factor in establishing person time at risk for drug exposures or drug therapy.

STRENGTHS

The UnitedHealth Group databases provide an efficient and unobtrusive method to determine and study exposures to prescription drugs. Further, drug

utilization and adverse events within UnitedHealth Group affiliated plans reflect practice in 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.⁶

The varied types of plan location and types of population, in combination with the large size of the databases, provide unique advantages to conduct pharmacoepidemiology research. Varied demographic characteristics of members provide strength for the utility of these databases when conducting this type of research. Specifically, populations of children (over 1 million), pregnant women (approximately 27 000 deliveries in 1997), and the elderly (almost 200 000 are 65 years of age or older) are sufficient for analysis given the large

size of the RDB. In addition, use of these standardized databases provides both numerators and denominators for exposures to drugs and allows the calculation of incidence and prevalence. Further, as an example below illustrates, rare exposures and rare outcomes (such as birth defects) can be detected given the size of the databases. Table 18.2 provides a listing of the top 25 drug exposures for health plan members for a 1 year time period in 1997–1998 and indicates the numbers of prescriptions and members available for analysis.

The databases also provide the ability to link various types of files longitudinally for individual members, regardless of the site of service as described above. Thus, adverse events and outcomes may be analyzed considering temporality in relation to exposure through pharmacy claims and linking such varied health services as hospitaliza-

Table 18.2. Top 25 outpatient prescription drugs from 12 UnitedHealth Group independent practice association (IPA) health plans between 1 April, 1997 and 31 March, 1998

Drug name	No. of prescriptions	No. of members
Estrogens, conjugated	824 499	168 234
Amoxicillin trihydrate	639 312	528 510
Ethinyl estradiol	639 259	162 666
Levothyroxine sodium	431 303	109 649
Albuterol sulfate	413 204	199 872
Fluoxetine hydrochloride	339 540	84 632
Lisinopril	334 024	69 960
Loratadine	307 210	128 361
Acetaminophen w/hydrocodone	291 151	168 855
Atenolol	257 552	60 161
Azithromycin dihydrate	254 100	207 118
Simvastatin	245 068	54 120
Omeprazole	238 745	69 753
Ranitidine hydrochloride	238 232	103 613
Insulin	229 920	45 067
Paroxetine hydrochloride	228 822	69 187
Hydrochlorothiazide w/triamterene	226 997	60 507
Trimethoprim	223 722	174 301
Estradiol	217 775	51 166
Cephalexin	207 085	177 773
Acetaminophen w/propoxyphene	204 511	109 262
Ibuprofen	204 363	136 868
Medroxyprogesterone acetate	203 733	59 647
Erythromycin	203 396	152 735
Beclomethasone dipropionate monohydrate	197 591	98 471

tion, emergency department use, physician visits, and any other site of health care service. Further, with respect to pharmacy claims, the prescribing physician can be determined as well as the specialty of that provider. Claims submissions are generally complete, since claims must be submitted by providers for payment at most of the UnitedHealth Group affiliated health plans.

UnitedHealth Group has conducted a number of studies that utilize other data sources in addition to administrative data, including medical records information and surveys. As Iezzoni⁷ suggests, 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 study¹ and additional information can then be obtained from medical records. This supplemental information can be crucial in pharmacoepidemiology studies. For example, using the abstraction of medical records one can confirm a diagnosis and obtain information on risk factors and outcomes. Using the above strengths of the administrative claims data, supplemented by other data sources, UnitedHealth Group has had experience conducting both cohort and case-control studies.

WEAKNESSES

Large databases must be used “judiciously”⁸ and require recognition of limitations as well as advantages. The constraints of a database must be recognized in order to fully understand and utilize data appropriately.⁹ For the UnitedHealth Group research databases, there are certain structural constraints that limit access to obtaining all possible prescription drug claims. Like many other resources for pharmacy management services, use of inpatient drugs is not available. In addition, given the pharmacy benefit structure if the cost of a prescription drug is lower than the co-payment 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. However, as the

list of exposures for the top 25 drugs suggests (Table 18.2), drug exposure is sizable.

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

With respect to the completeness of the databases, certain plans that have different financial incentives than 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 disadvantage is certain variables that would be of interest are not available on the electronic claim databases, such as race/ethnicity or information that a member is deceased. If necessary for a specific study, this information may be determined through review of the medical record. However, since medical records are not standardized, it is also possible that this information still may not be available.

Another limitation is the claims lag or the length of time required to obtain all claims for a given time frame. The claims lag is short for pharmacy claims (one month) but is longer for physician and facility claims (out to approximately 6 months). The claims lag may be variable across study years and should be taken into account in the design of a study whenever feasible.

In summary, certain disadvantages of using the UnitedHealth Group administrative claims databases may be taken into account or minimized through study design or the use of proxy measures. Others must simply be noted in any summary of study findings. The following examples provide empirical information on both the strengths and limitations of UnitedHealth Group data and present specific applications of the utilization of the databases to address pharmacoepidemiology questions.

PARTICULAR APPLICATIONS

The first summary of an application of the RDB is a feasibility study of the use of an osteoporosis medication in association with esophageal and gastric events. Two large scale studies are then discussed: (i) the association of an antibiotic taken during pregnancy with birth defects (including mother–baby sets), and (ii) a study of antibiotic associated diarrhea in a large population. Both of these studies involved the use of administrative data as well as medical records.

FEASIBILITY STUDY: OSTEOPOROSIS MEDICATION AND ESOPHAGEAL EVENTS

UnitedHealth Group conducts feasibility studies to estimate the rate of adverse events in the general population for specific events reported to FDA under the Spontaneous Reporting System. These evaluations are conducted using the administrative claims data from the RDB for 12 health plans affiliated with UnitedHealth Group, as described above. Preliminary descriptive data are obtained with a rapid turnaround time and usually involve drugs that have been newly introduced to the market. An example of such a feasibility study includes the incidence of adverse esophageal or gastric events in association with the use of a drug for osteoporosis, alendronate (Fosamax[®], Merck & Co., Inc.).

Alendronate was approved for use by the FDA in September 1995. Subsequently, case reports of esophagitis, esophageal ulcer, and esophageal stricture were published.^{10–12} In 1996, the manufacturer revised the product safety information to warn about this risk. UnitedHealth Group computerized pharmacy claims were searched to identify the 1321 members who received a prescription for alendronate between 1 January, 1996 and 30 August, 1996. The medical claims data for these alendronate users were then searched for subsequent diagnosis codes indicative of adverse esophageal and gastric events including esophagitis, ulcer of the esophagus, esophageal perforation, gastric ulcer, and gastritis/duodenitis. For the 1108 female and 102 male users with no esophageal or gastric diagnoses

prior to alendronate use, an incident diagnosis code of interest was identified for 39 users, 95% of whom were female. Thirteen of these cases had an endoscopic examination performed and five cases were hospitalized. The cumulative incidence of total upper gastrointestinal events was 3.3% in females, 2.0% in males, and 3.2% overall.¹³

STUDY OF AN ANTIBIOTIC AND ASSOCIATED BIRTH DEFECTS DURING PREGNANCY

A study was conducted to evaluate the potential for human teratogenicity for the antibiotic, clarithromycin (Biaxin[®], Abbott Laboratories), when used by women during pregnancy. Because previous findings from animal studies raised the question of an increased risk, clarithromycin was labeled a class C drug by the FDA and, therefore, is not recommended for use during pregnancy.¹⁴ The purpose of this study was to assess the rate and types of birth defect in infants born to women following exposure to clarithromycin in the first trimester of their pregnancy.

Claims data from UnitedHealth Group affiliated health plans were used to determine exposure to clarithromycin from 1991 to 1995. In 1991, the underlying population that was screened for exposure consisted of 1.2 million health plan members. The number increased each year with about 2.3 million health plan members screened in 1995. Pharmacy claims were used to identify women exposed to clarithromycin and hospital claims were used to identify women who had a delivery. Enrollment files were used to link the mother–infant sets. For each mother–infant set identified as exposed, medical records were abstracted to determine whether there was a birth defect outcome, since birth defect diagnoses from claims data at the time of delivery may not be complete.

Of the 48 003 women of childbearing age identified from pharmacy claims as having a prescription fill date for clarithromycin during the study period, 5372 also had a hospital claim for a delivery. Of these, the 315 women who had a clarithromycin fill date during the 270 days prior

to their delivery date were identified as potentially exposed during pregnancy (Table 18.3). Women were identified as potentially exposed during the first or second trimester using the assumptions that the first trimester occurred between 180 and 270 days prior to the delivery date for a full-term pregnancy and between 90 and 180 days to account for a possible preterm delivery. For efficiency in obtaining primary medical records, the eight health plans with the largest number of estimated exposed women were selected for the study. The regional representation for the eight health plans included five health plans from the Midwest, and one each from the Northeast, the Southeast, and the West. Of these 209 women, medical record abstractions were completed for 199 mother–infant sets, for a completion rate of 96%. A new calculation of the exposure period was made using the gestational age obtained from the medical record, which identified the 143 women and their 149 infants who were exposed to clarithromycin during the first trimester (Table 18.3).

The results showed that 91 (64%) of the exposed women had a prescription fill date during the first month of pregnancy, consistent with the likelihood that the drug was prescribed prior to the knowledge that the woman was pregnant. Birth malformations were identified for eight (all singleton births) of the 149 infants, excluding infants with undescended testicles likely to resolve ($n = 4$), for an observed overall birth defect rate of 5.8% (rate calculated with twins excluded for comparison). Of interest is that this rate was not statistically

significantly different compared to the expected rate of 4.5% based on national data.¹⁵ Of the five malformations considered major, there was no consistency in the types of malformation. There were two heart defects and one each of a chromosomal, a gastrointestinal, and a genital abnormality. Although study size precluded the detection of weak associations or an association with a specific outcome, the results suggested that clarithromycin was unlikely to be a major teratogen in humans.

This study provides an example of an efficient method to ascertain the potential for adverse outcome(s) where both the exposure and the outcome are rare. A subsequent analytic study could have been completed if these results had indicated a major problem. These capabilities can be used to address public health concerns or to evaluate therapeutic outcomes. In this situation, a claims based study supplemented by medical records provided a way to obtain human outcome information from a population of pregnant women who have been excluded from premarketing clinical drug trials.

STUDY OF ANTIBIOTIC-ASSOCIATED DIARRHEA

The purpose of this retrospective cohort study was to determine the relative risk of diarrhea and *Clostridium difficile* associated diarrhea for oral antibiotics dispensed in the ambulatory setting. Diarrhea is a common, although usually mild, adverse effect of antibiotics occurring in up

Table 18.3. Use of claims and medical record data in the process to identify women who were exposed to clarithromycin in the first trimester of pregnancy

Number	Description
149 001	Members with a pharmacy claim for clarithromycin
48 003	Women who were 14–44 years of age
5 372	Women who had a delivery
315	Women who were identified as potentially exposed to clarithromycin during pregnancy
235	Women who were identified as potentially exposed to clarithromycin during the first or second trimester
209	Women (and 217 infants) who were from the eight largest health plans
199	Mother–infant sets for whom medical records were obtained
143	Women (and 149 infants) who were identified as exposed to clarithromycin in the first trimester

to 5–30% of individuals receiving antibiotic therapy.^{16–19} A small proportion of patients develop severe antibiotic-associated diarrhea with evidence of colitis and pseudomembranous colitis requiring hospitalization.¹⁶ Factors that may influence the risk of diarrhea include type of antibiotic, exposure to multiple antibiotics, duration of antibiotic use, young or elderly age, pre-existing gastrointestinal problems, abdominal surgery, and care provided in an institutional setting.^{16–19} Antibiotic use has been consistently associated with an increased risk of infection with *C. difficile*.^{19–21}

The study included 358 389 members enrolled in four UnitedHealth Group affiliated health plans between 1 July, 1992 and 31 December, 1994, located in the Southeastern, Northeastern, and Midwestern United States. Members were followed for up to 42 days following the date of each dispensed antibiotic. Survival analysis was used to estimate the risk of diarrhea accounting for the number of days at risk for a single antibiotic group or for the first antibiotic in multiple antibiotic group risk periods. The analyses were adjusted for the potentially confounding effects of age, gender, health plan, and selected medical conditions obtained from claims data. Results indicated 3230 events of diarrhea (90/10 000 risk periods) determined by claims data with a diagnosis for diarrhea. Six different antibiotics (including two quinolone antibiotics, ofloxacin, and ciprofloxacin) had significantly higher risk of diarrhea in adults 21 to 50 years of age when compared to amoxicillin (Table 18.4).²² A separate analysis of the cephalosporins showed that, compared to first

generation cephalosporins, the adjusted relative risk of diarrhea was greater for second generation cephalosporins (RR = 1.45, $p < 0.01$) and, although not statistically significant, was greater for a third generation cephalosporin.

The results of this study showed that several antibiotics had greater risk for diarrhea compared to amoxicillin. Limitations of this study include the potential underidentification of diarrhea using diagnosis codes, and the inability to identify type or severity of diarrhea from claims data alone. A strength of this study was the large sample size across multiple antibiotic groups allowing for calculation of relative risks after adjusting for potential confounding factors. Use of population based data from multiple geographically diverse health plans can contribute to the understanding of risks of drugs available in the ambulatory setting in association with adverse events.

THE FUTURE

Postmarketing surveillance and pharmacoepidemiology studies are, by necessity, conducted in the context of a rapidly changing health care environment. Looking to the future this dynamic environment has implications for the nature of the data obtained, the ability to obtain information, and characteristics of the general and specific populations that form the basis of these studies. Changes in the health care system, given the concern with the cost of health care, will continue to impact the health benefits structure, including pharmacy benefits coverage. As the pharmacy benefit struc-

Table 18.4. Antibiotics which had a statistically significant elevated risk of diarrhea compared to the amoxicillin reference group

Antibiotic	Risk periods (21–50 years of age)	% of risk period with diarrhea	Relative risk (p value)
Amoxicillin	45 506	0.43	1.0
Cefixime	1 202	0.92	2.08 ($p < 0.05$)
Cefaclor	2 570	0.82	1.73 ($p < 0.05$)
Cefuroxime	8 316	0.90	1.97 ($p < 0.001$)
Ciprofloxacin	5 693	2.20	3.17 ($p < 0.001$)
Amoxicillin/clavulanic acid	8 673	0.76	1.64 ($p < 0.001$)
Ofloxacin	2 511	1.43	2.91 ($p < 0.001$)

ture is broadened to include coverage across a larger number of defined tiers (e.g. brand, generic, and not on the preferred drug list), certain transactions may not be captured in the pharmacy files if the cost is lower than the copayment. This may, however, be balanced by the increasing cost of prescription drugs.

The political environment reflects increasing sensitivity by the public to use of medical information and medical records (see also Chapter 26). Reflecting this, Congress required in passage of the Health Insurance Portability and Accountability Act (HIPAA), that Federal confidentiality standards for medical information be adopted by Congress or by the Department of Health and Human Services in 1999.²³ In a study conducted in a state with stringent regulations (Minnesota), the ability to complete medical record abstraction was reduced to a completion rate of approximately 20% when study-specific informed consent was required.²⁴ When such a low percentage of the study population is willing to participate, this may interject potential but unknown bias into the results. Regulations at both the Federal and State levels will potentially compromise the ability to conduct health care research of any type, if they are not carefully balanced to allow access to medical records for bona fide research purposes while also protecting patient confidentiality. International concern has also arisen, evidenced by the promulgation of the European Privacy Directive and the International Society of Epidemiology (ISPE) professional association guidelines for data privacy.²⁵

With respect to the ability to obtain data, additional automation may be available in the future for health care companies such as UnitedHealth Group. Laboratory test results and medical records may become available electronically. However, the process is likely to be stymied both by a lack of standards and by the potential high cost to implement electronic medical records across a diverse network of independent practitioners. The size of the research databases will increase in the future as other UnitedHealth Group affiliated health plans are added to the RDB using commonly defined data elements.

The importance of pharmacoepidemiology and the ability to conduct postmarketing surveillance will become increasingly critical in the future due to a number of factors relating to changing characteristics of the American population. First, changes attributable to the aging of the US population will augment the need for a better understanding of the use of prescribed medications in the older population, including variations in metabolism and appropriate dosages. Second, as the population ages, more Americans are likely to receive a larger number of medications due to the prevalence of chronic disease in this group, raising polypharmacy questions. Increased use of over-the-counter (OTC) drugs, alternative medications, and devices will also need to be explored. Third, developments in the biotechnology field may also result in new formulations of drugs that will need to be addressed by reviewing adverse events related to prescription medications in groups with different biological factors. Finally, as a larger number of drugs are approved by FDA in a faster timeframe, given the acceleration of technological innovation,²⁶ the potential for adverse events in both the general population and the elderly population will increase.

As innovation accelerates, postmarketing surveillance will become more crucial as society attempts to balance public health concerns with individual access to new therapies. The challenge for the future will be in balancing these competing demands. The information gained from the results of pharmacoepidemiology research using data from national health care settings will help meet this challenge.

REFERENCES

1. Selby JV. Linking automated databases for research in managed care settings. *Ann Intern Med* 1997; **127**: 719–24.
2. Quam L, Ellis LBM, Venus P, *et al.* 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.
3. Heinen L, Peterson E, Pion K, Leatherman S. Quality evaluation in a managed care system:

- comparative data to assess health plan performance. *Managed Care Q* 1993; **1**: 62–76.
4. Leatherman S, Chase D. Using report cards to grade health plan quality. *J Am Health Policy* 1994; 32–40.
 5. Peterson E, Shatin D, McCarthy D. Health services research at United HealthCare Corporation: the role of the Center for Health Care Policy and Evaluation. *Med Care Res Rev* 1996; **53**: S65–S76.
 6. Brook RH, Lohr KN. Efficacy, effectiveness, variations, and quality. Boundary-crossing research. *Med Care*, 1985; **23**: 710–22.
 7. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997; **127**: 666–74.
 8. Hui SL. Measuring quality, outcomes, and cost of care using large databases. *Ann Intern Med* 1997; **127**: 665.
 9. Wray NP, Ashton CM, Kuykendall DH, Hollingsworth JC. Using administrative databases to evaluate the quality of medical care: a conceptual framework. *Soc Sci Med* 1995; **40**: 1707–15.
 10. Maconi G, Porro GB. Multiple ulcerative esophagitis caused by alendronate. *Am J Gastroenterol* 1995; **90**: 1889–90.
 11. Sorrentino D, Trevisi A, Bernardis V, DeBiase F, Bartoli E. Esophageal ulceration due to alendronate. *Endoscopy* 1996; **28**: 529.
 12. Abdelmalek MF, Douglas DD. Alendronate-induced ulcerative esophagitis. *Am J Gastroenterol* 1996; **91**: 1282–3.
 13. Park B-J, Clouse J, Shatin D, *et al.* Incidence of adverse esophageal and gastric events in alendronate users. *Pharmacoepidemiol Drug Safety*. 51 In press.
 14. Prescribing information for clarithromycin. *Physicians' Desk Reference*®, 50th edn. Montvale, NJ: Medical Economics, 1996; 405–10.
 15. Heinonen OP, Slone D, Shapiro S. *Birth Defects and Drugs in Pregnancy*. Littleton, MA: Publishing Sciences, 1997.
 16. Ewe K. Diarrhoea and constipation. *Baillieres Clin Gastroenterol* 1988; **2**: 353–84.
 17. Hooker KD, DiPiro JT. Effect of antimicrobial therapy on bowel flora. *Clin Pharm* 1988; **7**: 878–88.
 18. George WL. Antimicrobial agent-associated colitis and diarrhea: Historical background and clinical aspects. *Rev Infect Dis* 1984; **6** (suppl 1): S208–13.
 19. Bartlett JG. Antibiotic-associated diarrhea. *Clin Infect Dis* 1992; **15**: 573–81.
 20. Fekety R, Shah AB. Diagnosis and treatment of *Clostridium difficile* colitis. *J Am Med Assoc* 1993; **269**: 71–5.
 21. Kelley CP, Pothoulakis C, LaMont JT. *Clostridium difficile* colitis. *New Engl J Med* 1994; **330**: 257–62.
 22. Clouse JC, Van Vorst K, Stergachis A, *et al.* A study of the risk of antibiotic-associated diarrhea and the risk of antibiotic-associated *Clostridium difficile* diarrhea identified in the ambulatory setting. *Pharmacoepidemiol Drug Safety* 1997; **6** (suppl 2): S67.
 23. Congressional Research Service 1998. *CRS Issue Brief: Medical Records Confidentiality*. Washington, DC: Library of Congress, 1998.
 24. McCarthy DB, Shatin D, Drinkard CR, *et al.* Medical records and privacy: empirical effects of legislation. *Health Serv Res* 1999; **34**: 417–25.
 25. Andrews E, Mann RD. ISPE position paper: Data privacy, medical record confidentiality and research in the interest of public health. *Pharmacoepidemiol Drug Safety*, 1999; **8**: 247–60.
 26. Post-1990 launches represent 43% of Rx market, IMS says. *The Pink Sheet: Prescription Pharmaceuticals and Biotechnology* 1998; **60**: 9–10.

19

Medicaid Databases

JEFFREY L. CARSON

University of Medicine and Dentistry of New Jersey—Robert Wood Johnson Medical School, New Brunswick, NJ,
USA

WAYNE A. RAY

Vanderbilt University School of Medicine, Nashville, TN, USA

BRIAN L. STROM

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

INTRODUCTION

The Medicaid system is the United States' health insurance system created in 1965 to provide access to medical care to some categories of economically disadvantaged and disabled persons. During the past two decades, Medicaid billing databases have been used for health services research and to examine the utilization and effects of drugs. In this chapter, we will explore the current status of these databases, and our view of the appropriate role of Medicaid data in pharmacoepidemiology research. We will first describe the Medicaid program in general. We will then review the three Medicaid databases that have most often been used for academic pharmacoepidemiology research: the Computerized On-Line Medical Pharmaceutical Analysis and Surveillance System (COMPASS[®]), the Tennessee Medicaid system, and the New Jersey Medicaid system. We will then explore the advantages and disadvantages of Medicaid systems, and provide examples of studies that have

been performed using this data. We will conclude with a description of the situations in which Medicaid data are most useful for pharmacoepidemiology research.

DESCRIPTION

MEDICAID PROGRAM

“The Medicaid Program” is a series of 54 programs, each supported jointly by federal and state funds and each managed independently by a US state or jurisdiction. Eligibility criteria are determined separately by each state, although certain groups must be covered in order to qualify for the federal contribution. Until the 1996 Welfare Reform Act, these included individuals covered under the Aid to Families with Dependent Children Program (AFDC), single parent families who met state determined income standards, and individuals eligible for the Supplemental Security

Income Program (SSI) because they are aged, blind, or disabled, and have limited income. Other needy patients also could be covered by the state, although federal matching funds were not available to pay for these benefits. Since the 1996 Welfare Reform Act, eligible individuals include children under age six and pregnant women whose family income is at or below 133% of the Federal poverty level; children born after 30 September, 1983 who are under age 19, in families with incomes at or below the federal poverty level; individuals eligible for the Supplemental Security Income Program (SSI) because they are aged, blind, or disabled, and have limited income; recipients of adoption or foster care assistance under Title IV of the Social Security Act; certain Medicare beneficiaries; and special protected groups, including people who lose their cash assistance due to income from work or from increased Social Security benefits.¹ Other needy patients may also be covered by the state for which federal matching funds are available to pay for these benefits. These optional groups are similar to the mandatory groups but eligibility criteria are more lenient. Included are others considered “categorically needy,” and also a group considered “medically needy,” who would be eligible for Medicaid under one of the mandatory or optional groups, except that their income and/or resources are above the eligibility level set by their state. The latter may qualify immediately, or may “spend down” by incurring medical expenses that reduce their income to or below their State’s medically needy income level. Because of these requirements, the population served by Medicaid differs significantly from the overall US population, with an overabundance of children, females, and non-whites (see Figure 19.1).

Medicaid reimburses providers through both fee-for-service and capitated models. A registry of persons eligible for Medicaid and records of services provided are maintained in computerized files. For data elements related to claims payment, there is routine auditing of data accuracy and completeness, resulting in high data quality.² However, because Medicaid programs are administered by the individual states, there is substantial variation between the states in enrollment

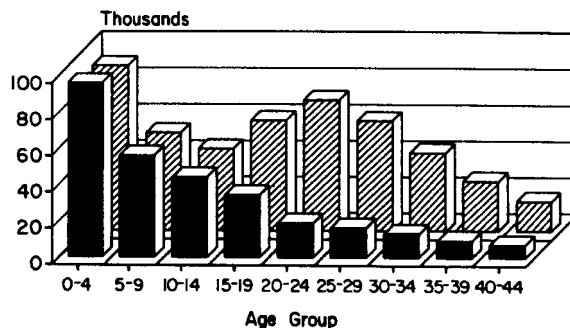


Figure 19.1. Age/sex distribution of Michigan Medicaid users, 1980–1986. Courtesy of Bud Gerstman, DVM, MPH, US Food and Drug Administration. $n = 767\,781$; $M = 295\,784$; $F = 471\,997$. ■ males; ▨ females.

criteria, program operation, and data element availability/quality.

The services provided to Medicaid recipients also vary among the states. The minimum services mandated by federal law include physician services, home health care services, nurse–midwife services, care in skilled nursing facilities, inpatient hospital care, outpatient hospital care, rural health clinic services, independent laboratory and radiology services, early and periodic screening of children, family planning services, and transportation to and from medical services. Many other services are also provided at the option of the state. Although not a mandatory service, virtually all states provide for reimbursement for prescribed drugs. The number of drugs eligible for reimbursement also varies among the states.

Research using Medicaid data from multiple states has been simplified by the development of the Medicaid Management Information System (MMIS), a set of specifications for computerized claims processing and management information. This was created in 1972 by the Department of Health, Education, and Welfare for fiscal and administrative control of this very large health care program. MMIS has six components: recipient files, provider files, claims processing files, reference file, surveillance and utilization review, and management and administrative reporting. Each state may tailor the system to their specific needs, although minimum standards must be met.

The recipient file includes a unique identifier for each patient in the database. If accurate, this file would be useful for determining the enrollment and followup status of cohorts of patients. This is especially important because of changes in the eligibility of Medicaid patients. In Tennessee, for example, at the end of one year, 25% lost enrollment and 3% had died, leaving 72% available for follow-up. After five years, 49% of the cohort were still alive and under observation.³ Loss of eligibility was greatest in children and young adults.³ The quality of the eligibility files varies among the different states. In particular, patients can sometimes become retroactively eligible or ineligible, sometimes for claims from years before.

In many states maternal and child records can be linked, thereby permitting investigations of birth defects that may be associated with drugs or procedures. For some states, this link is inherent in the individuals' identification numbers. In other states, separate linking files are needed.

The claims processing files include information on age, sex, state, inpatient and outpatient diagnoses (by ICD-9-CM code), outpatient drugs (by NDC code), deaths (in some states), and procedures, such as laboratory and radiographic procedures. The pharmacy data consists of records of all outpatient and nursing home prescriptions filled at the pharmacy for drugs and medical equipment/supplies included on the Medicaid formulary. There is generally one claim for each filled prescription and for each refill. This record contains date the prescription was filled, a code identifying the drug dispensed (currently the National Drug Code), quantity of the drug dispensed, including the days the supply of drug is anticipated to last and the ID of the pharmacy and prescribing physician. Usually no more than a 30 day supply of medication can be dispensed at one time, although some states permit or even encourage a larger supply (e.g., 90 days) to be dispensed for chronically used drugs.

The diagnosis data contain the records of hospitalizations and outpatient claims for care provided to Medicaid enrollees. For hospitalizations, information in this file includes the identity of the hospital, and can include the hospital

admission and discharge dates. For non-Medicare recipients, this file can also contain primary and secondary diagnoses and surgical procedures. For enrollees over 65 years of age, Medicare is usually the primary payer. The resulting Medicaid claims ("crossovers") may not retain complete information. Medicaid files can be linked with Medicare Part A files to obtain the missing data elements, which include admission diagnosis, up to five discharge diagnoses, and surgical procedures billed. In the Medicare inpatient files, both diagnoses and procedures are coded with the ICD-9-CM system.

Other Medicaid encounter files include nursing home, home health care (which can be augmented with Medicare data), ambulance services, vision and dental services, durable medical equipment, and other miscellaneous services.

COMPUTERIZED ON-LINE MEDICAL PHARMACEUTICAL ANALYSIS AND SURVEILLANCE SYSTEM[®]

The Computerized On-Line Medical Pharmaceutical Analysis and Surveillance System (COMPASS[®]) was originally conceived by Health Information Designs, Inc. (HID), and it has been developed by HID, in part under contract to the FDA, since 1977.^{4,5} This database is currently owned by Protocare Sciences, Herndon, VA, and no longer called COMPASS[®], but rather the Protocare Sciences Proprietary Medicaid Database. However, most, if not all, published papers have come from the older COMPASS[®] database. Researchers at the University of Pennsylvania School of Medicine and the UMDNJ–Robert Wood Johnson Medical School have assisted this development since 1980. The original database included billing data from Medicaid patients in the following states: Florida, Missouri, Nebraska, Arkansas, Mississippi, Ohio, Michigan, Virginia, Colorado, Maryland, and Minnesota, bringing the total Medicaid population to over eight million patients. However, over the past 3 years, new data are only available for approximately 1.25 million patients from Ohio. There is up to a 2-month lag period between the time a drug is dispensed and the time that it appears on the system. Diagnoses may lag up to three

DATE: 08/17/90
PAGE: 0003

STATE 4 MEDICAID PROGRAM
DURBASE - NEBRASKA
PATIENT DX/RX HISTORY PROFILE

HEALTH STATE 4
INFORMATION & PHARMACISTS
DESIGNS, INC. ASSOCIATION

PATIENT ID: SEX: F AGE: 58 (1989) COUNTY: 40 RACE: 1

DATE	DIAGNOSIS	DRUG	STRENGTH	QTY & DAY	PHYS/HOSP	PHARMACY	MS
85053	EPISTAXIS	7847			I-45023347004		
85053	CHRONIC RENAL FAILURE	585			D-50862915000	47013710000	
85058		00651102	HYDRALAZINE TABLET 50 MG (C) (MAC) .038	100			
25065		00690502	LOPRESSOR TABLET 100 MG	60			
85076	EPISTAXIS	7847		1	D-50862915000	47013710000	
85076	EPISTAXIS	7847			O-45023347004		
85077	CHRONIC RENAL FAILURE	585			P-47059252300		
85077	EPISTAXIS	7847			I-45023347004		
85077	HYPERTENSION NOS	4019			I-45023347004		
85078	CARDIOMEGALY	4293			P-47058719903		
85081	PLEURAL EFFUSION NOS	5119			P-47058719903		
85084	CONGESTIVE HEART FAILURE	4280			P-47058719903		
85089		00453512	CHLOR-TRIMETON ALLERGY TABLET	20	D-47064618100	47013710000	
85089		00560902	SULFAMETHOXAZOLE 800/TRIMETHOPRIM-DS	20	D-47064618100	47013710000	
85092	ABDOMINAL PAIN	7890			P-47058719901		
85092	SPRAIN THORACIC REGION	8471			I-45023347004		
85092	PLEURAL EFFUSION NOS	5119			P-47058719901		
85092	CONGESTIVE HEART FAILURE	4280			P-47059252300		
85092	CONGESTIVE HEART FAILURE	4280			I-45023347004		
85093	HEART DISEASE NOS	4299			P-47058719901		
85093	BONE CARTILAGE DIS NEC	73399			P-47058719901		
85096		00623412	LASIX TABLET 80 MG	60	D-47039358706	47013710000	
85096		00741410	ALUPENT INHALANT SOLUTION, 5 PERCENT, 10	1	D-50862915000	47013710000	
85096	CONGESTIVE HEART FAILURE	4280			P-47059252300		
85100	OTITIS MEDIA NOS	3829			P-47065086213		
85105		00434601	BASALJEL CAPSULE	100	D-50862915000	47013710000	
85105		00651102	HYDRALAZINE TABLET 50 MG (C) (MAC) .038	100	D-50862915000	47013710000	
85105		00623412	LASIX TABLET 80 MG	60	D-47039358706	47013710000	
85110	ASTHMA W/O STATUS ASTHM	49390			O-45023347004		
85112		00370077	PERCOCET-5 TABLET 5MG	100	D-50862915000	47013710000	
85113		00741501	VENTOLIN INHALER 17 GRAM	1	D-50862915000	47013710000	
85117	CHRONIC RENAL FAILURE	585			I-45023347004		
85117	PULM CONGEST/HYPOSTASIS	514			P-47058719901		
85117	CONGESTIVE HEART FAILURE	4280			I-45023347004		
85117		00340677	VALIUM TABLET, 5 MG (500 PRICE)	50	D-50862915000	47013710000	
85117		00623412	LASIX TABLET 80 MG	60	D-47039358706	47013710000	
85119	PLEURAL EFFUSION NOS	5119			P-47058719901		
85120	CHRONIC RENAL FAILURE	585			P-47037653104		
35120	HYPERTENS RENAL DIS NOS	4039			I-47037653101		
35120	RENAL FAILURE NOS	586			P-47052846412		
85120	CHRONIC RENAL FAILURE	585			I-47037653101		
85122	RENAL FAILURE NOS	586			P-47063668713		
35124	ADJUSTMENT REACTION NOS	3099			P-47062996700		
35131	CHRONIC RENAL FAILURE	585			O-47046185901		
35131	ELECTROLYT/FLUID DIS NEC	2769			D-47039358706	47013710000	
35133		00522205	CYPROHEPTADINE TABLET 4 MG	50	P-47059252300		
35137	OSTEOARTHROS NOS-PELVIS	71595					

Figure 19.2. Sample COMPASS® profile.

months. The currently available billing data for each patient are identical to those previously described. Importantly, because of the increasing societal concern about confidentiality (see Chapter 26), medical record access is no longer available.

An example of a patient profile from the Ohio Medicaid database is provided in Figure 19.2. Listings of its most common drugs and diagnoses appear in Tables 19.1 and 19.2. More detailed descriptions of the system have been published elsewhere.⁴⁻⁶

A number of studies have now been published using the system.⁷⁻¹⁴ Some are described later in this chapter.

TENNESSEE MEDICAID DATABASE

Since 1973, the Department of Preventive Medicine at Vanderbilt University School of Medicine has had a contract from the state of Tennessee to collect the claims files of the Medicaid program, to analyze the quality of the claims, and to perform research tasks of interest to the state. In 1997, there were 1 421 963 enrollees, 27% of the state's population.¹⁵ There were 156 657 enrollees 65 years of age and older and 31 910 were 85 years of age and older. In 1996, there were 35 707 births to women enrolled in Tennessee Medicaid, which accounted for 49.4% of births in Tennessee.

Prior to January 1994, the Medicaid program in Tennessee was a vendor payment, fee-for-service program. It then changed to a capitated model called TennCare.^{16,17} The program licenses 12 managed care organizations and requires each enrollee to select an managed care organization. The managed care organizations receive monthly capitation payments and each has its own formularies that restrict the reimbursable drugs. The quality of the database has generally been maintained. However, further work to investigate the accuracy of outpatient claims is in progress, because there are no longer financial incentives for complete and accurate data.

The Tennessee system has developed linkages to files that include information not available in Medicaid. These include the following.

1. *Medicare files to Medicaid enrollment.* This provides a consistent record of inpatient admissions, outpatient visits and short nursing home stays for persons 65 and older.
2. *Vital statistics files to Medicaid enrollment.* Birth and death certificate files are obtained annually from the Tennessee State Center for Health Statistics. The birth certificate file establishes Tennessee residence and provides additional information such as birthweight, gestational age, and parity. The death certificate file contains date of death and underlying cause of death.
3. *Medicaid mothers and children linkage.* This link permits study of fetal exposures because of the capacity to identify pregnancies, establish their duration (gestational age), and identify the resultant child, thus enabling maternal/fetal drug effects studies.
4. *Public health clinic files.* These are used to link information on routine childhood immunizations administered at public health clinics permitting the study of vaccine safety.
5. *Motor vehicle files.* The Medicaid enrollment file is linked with the Tennessee Department of Transportation's file of licensed drivers and motor vehicle crash reports for the period 1984 to 1988. This database was created to study the effects of medication use among the elderly on driving safety.^{18,19}
6. *Tennessee cancer registry.*

Researchers using the Tennessee Medicaid database currently have access to medical records. For each study, the procedure involves obtaining permission from the Medicaid program and then contacting each provider in advance to explain the specific study for which records are requested. Trained nurse-abstractors then travel to the facility and complete structured study forms. With this procedure, between 80 and 95% of hospital medical records identified in the computer are obtained.²⁰⁻²²

Medicaid enrollment and encounter files are updated on a quarterly basis. Pharmacy data and data for the elderly from Medicare are reasonably current. Inpatient and outpatient encounter files for persons <65 years of age may be delayed by

Table 19.1. Most frequent diagnoses—January 1995–June 1997 Ohio Medicaid data*

ICD code	Diagnosis	Females	Males	Total
4659	Acute URI Nos	277 465	195 884	473 349
V202	Routine Child Health Exam	214 374	217 099	431 473
3829	Otitis Media Nos	180 519	154 582	335 101
2398	Neoplasm Nos, Site Nec	224 321	105 350	329 671
462	Acute Pharyngitis	196 971	120 412	317 383
V726	Laboratory Examination	220 589	80 396	300 985
V725	Radiological Exam Nec	174 534	84 728	259 262
4660	Acute Bronchitis	140 384	81 022	221 406
5990	Urinary Tract Infection Nos	173 281	30 171	203 452
7890	Abdominal Pain	145 778	52 021	197 799
78650	Chest Pain Nos	126 633	67 739	194 372
5589	Noninfectious Gastroenteritis Nec	111 909	76 876	188 785
3679	Refraction Disorder Nos	119 350	61 471	180 821
490	Bronchitis Nos	105 877	63 038	168 915
78609	Respiratory Abnormal Nec	103 736	64 061	167 797
7806	Pyrexia Unknown Origin	84 162	70 278	154 440
4019	Hypertension Nos	98 599	44 893	143 492
6929	Dermatitis Nos	86 058	57 111	143 169
7840	Headache	98 558	39 457	138 015
486	Pneumonia, Organism Nos	73 003	54 834	127 837
4739	Chronic Sinusitis Nos	82 535	42 101	124 636
V723	Gynecologic Examination	121 905	586	122 491
4619	Acute Sinusitis Nos	82 220	40 031	122 251
7862	Cough	72 669	47 381	120 050
V222	Preg State, Incidental	109 420	171	109 591
7295	Pain In Limb	71 447	35 794	107 241
49390	Asthma W/o Status Asthma	63 409	42 345	105 754
3671	Myopia	70 341	33 384	103 725
1101	Dermatophytosis of Nail	68 230	29 166	97 396
3670	Hypermetropia	56 862	35 108	91 970
61610	Vaginitis Nos	91 170	565	91 735
7242	Lumbago	62 816	28 159	90 975
V270	Deliver-Single Liveborn	90 450	85	90 535
463	Acute Tonsillitis	50 861	37 034	87 895
37230	Conjunctivitis Nos	47 502	37 711	85 213
V221	Supervised Otherwise Normal Pregnancy	84 517	211	84 728
0799	Viral Infection Nos	47 830	36 533	84 363
25000	Diabetes Uncomplicated Adult	59 792	24 051	83 843
2859	Anemia Nos	55 917	27 476	83 393
920	Contusion Face/Scalp/Neck	42 250	36 502	78 752
311	Depressive Disorder Nec	57 640	20 648	78 288
7245	Backache Nos	53 530	22 584	76 114
496	Chronic Airway Obstruct Nec	45 725	29 611	75 336
4779	Allergic Rhinitis Nos	44 985	28 370	73 355
V270	Eye Vision Examination	48 399	24 950	73 349
0340	Strep Sore Throat	42 871	30 436	73 307
650	Normal Delivery	72 996	80	73 076
2765	Hypovolemia	45 025	26 855	71 880
7870	Nausea And Vomiting	45 674	24 462	70 136
7807	Malaise And Fatigue	50 894	18 456	69 350

*Represents unduplicated counts of diagnoses for 1 921 006 patients.

Table 19.2. Most frequently used drugs—January 1995–June 1997 Ohio Medicaid data*

Drug	Females	Males	Total
Acetaminophen W/Codeine	97 276	39 278	136 554
Amoxil (250 mg/5 ml)	63 095	63 182	126 277
Trimox 250 (250 mg/5 ml)	59 132	58 838	117 970
Cipro (500 mg)	81 689	30 907	112 596
Biaxin	78 024	30 817	108 841
Trimox 500	71 259	25 543	96 802
Cardec-DM	47 733	38 614	86 347
Proventil	56 825	28 412	85 237
Amoxil (125 mg/5 ml)	41 408	43 070	84 478
Propoxyphene Napsylate W/Apap	59 882	19 750	79 632
Nix	50 984	28 622	79 606
Trimox 125 (125 mg/5 ml)	36 386	37 058	73 444
Hydrocodone W/Acetaminophen	49 636	22 020	71 656
Bactroban	38 478	26 971	65 449
Zithromax	50 014	15 148	65 162
Trimox 250	45 789	19 106	64 895
Ultram	46 566	17 618	64 184
Zantac	44 685	18 946	63 631
Zoloft	47 519	16 079	63 598
Toradol	44 657	17 418	62 075
Albuterol Sulfate	26 200	33 249	59 449
Acetaminophen W/Codeine	42 478	16 597	59 075
Claritin	42 075	15 281	57 356
Macrobid	52 325	3 380	55 705
Amoxicillin	40 796	14 401	55 197
Ventolin	35 887	19 192	55 079
Terazol 7	54 628	317	54 945
Sulfatrim Pediatric	30 961	23 431	54 392
Veetids 500	37 642	16 299	53 941
Prozac	39 051	13 360	52 411
Axid	36 414	15 617	52 031
Lotrisone	34 529	16 855	51 384
Cardec-DM	24 421	25 644	50 065
Augmentin 500	33 215	15 801	49 016
Poly-Histine DM	27 596	20 900	48 496
Augmentin 250	32 359	15 349	47 708
Ceftin	34 928	12 745	47 673
Lotrisone	30 393	16 755	47 148
Phenergan	35 999	10 180	46 179
Paxil	34 416	11 298	45 714
Augmentin 250 (250 mg/5 ml)	22 159	23 213	45 372
Daypro	32 560	11 550	44 110
Zantac	30 631	13 156	43 787
Prilosec	30 692	12 870	43 562
Duricef	29 037	14 219	43 256
Ery-Tab	32 070	11 119	43 189
Nitrostat	29 399	13 784	43 183
Cipro (250 mg)	33 148	9 325	42 473
Zithromax	31 342	9 289	40 631
Albuterol	27 179	13 436	40 615

* Represents unduplicated counts of drugs for 1 921 006 patients.

medical care organizations' late submission of data to the state. The vital records files are available approximately 10 months after the end of a calendar year.

Early in its use, the Tennessee database was principally used for drug utilization studies.²³⁻²⁸ For example, studies have examined the use of tetracycline and chloramphenicol and demonstrated the frequent inappropriate use. Efforts to change prescribing behavior with educational interventions were then demonstrated to be effective.

More recently, several important analytic epidemiology studies have been performed: psychotropic drugs were demonstrated to be associated with hip fractures,²⁹ nonsteroidal anti-inflammatory drugs were demonstrated to be associated with fatal peptic ulcer and upper gastrointestinal bleeding,^{30,31} corticosteroids were demonstrated to be associated with peptic ulcer disease when taken with nonsteroidal anti-inflammatory drugs,³² childhood immunizations were demonstrated to be associated with seizures,³³ hip fractures were demonstrated to occur frequently during the first week after discharge from a hospital,³⁴ risk factors for hypoglycemia were demonstrated,³⁵ and association between hypoglycemia and antihypertensive drugs.³⁶

NEW JERSEY MEDICAID DATABASE

Avorn and associates have obtained access to Medicaid data from New Jersey, including data on approximately 700 000 patients from 1980 to 1996. These data sets are linked with Medicare data. In addition, these data can be combined with drug exposure information from the 250 000 recipients 65 years of age or greater enrolled in the Pharmacy Assistance Aging Program. No mechanism for accessing primary medical records has yet been established and tested. Examples of how these data have been used include examination of the effect of changes in reimbursement for drugs and its effect on drug utilization,³⁷ the frequency of nursing home placement,³⁸ the use of antidepressants in patients prescribed β -blockers,³⁹ and the association between antihypertensive drug therapy and initiation of medications for diabetes mellitus.⁴⁰

STRENGTHS

The most important advantage of Medicaid databases is their size. Currently, the three states actively in use by pharmacoepidemiology investigators each have between 1 and 2 million patients available for study. This very large sample size permits studies to be performed even when the drug exposure or the disease of interest is uncommon. A related advantage is that this sample of study subjects can be examined at a relatively small cost, since the data are generated as a by-product of a billing and administrative system. While using such a system is not inexpensive, a study of an equal number of patients that collected data *de novo* would be orders of magnitude more costly. In addition, because the data already exist, studies can be carried out relatively quickly.

Medicaid databases ostensibly include all reimbursed medical care provided by any provider to eligible patients, although it remains to be established whether Medicaid managed care has changed that. Thus, the system includes information on medical care that the patient may not remember and any given provider may not be aware of. In addition, both inpatient and outpatient diseases can be studied. Also, the system is, theoretically, population based, which permits the calculation of incidence rates. In reality, because of difficulties defining the eligible population at any one time due to continually changing eligibility, the population that sometimes has been used for these calculations consists of all those with a claim for medical services during a defined time frame.

A very important advantage of Medicaid data is that the data on exposure are not subject to recall or interviewer bias. Recall bias occurs when groups of patients who are being compared to each other differ in their ability to recall antecedent exposures or events. For example, this might represent a special problem in a case-control study of drug-induced birth defects,⁴¹ in which a mother of an abnormal child may be much more likely to remember prior drug exposures than a mother of a normal child. Interviewer bias occurs when groups of patients who are being compared to each other differ in the way they are questioned by the

study's interviewers. This could occur, for example, if a patient with an outcome of interest is questioned more intensively about prior drug exposure than a patient without the disease. Since drug exposure is defined by a claim by a pharmacist for reimbursement and is collected before the outcomes occur, neither of these potential biases are a problem in a billing database.

There is now agreement that automated pharmacy claims are one of the best sources of information on drug use, although this source also has its limitations. The primary issues are patient compliance and the use of drugs from other sources. Several studies have found excellent concordance between pharmacy records and patient self-reports.¹²²⁻¹²⁴ Data from Tennessee nursing homes have found excellent concordance between medical administration records and the Medicaid prescription files (unpublished data). The major problem areas are drugs taken intermittently for symptom relief, over-the-counter drugs, drugs not on the formulary, drugs usually given by injection, immunizations, and drugs given in a hospital.

An advantage of Medicaid databases, even relative to other automated databases, is the overrepresentation of special populations. Medicaid has substantially greater numbers of pregnant women, young children, the elderly, nursing home residents, and African Americans than would be expected in a population of comparable size. Because these populations often are excluded from or underrepresented in premarketing trials, it is particularly important to include them in post-marketing studies. Published studies for these vulnerable populations include fetal metronidazole exposure and childhood cancer,⁴² angiotensin converting enzyme inhibitors and sharply increased risk of angioedema in African Americans,⁴³ and safety of β -blockers in elderly users of hypoglycemics.³⁶

WEAKNESSES

GENERALIZABILITY

Because of the skewed nature of the population served by Medicaid, one must be concerned that

results obtained with these databases may not be generalizable to non-Medicaid patients. This is less of a problem for analytic studies, i.e., cohort and case-control studies, as both their study group and their control group come from the same population. In fact, it can be considered advantageous, as it may help to control for socioeconomic status as a potential confounding variable. As an example, studies evaluating the gastrointestinal side effects of nonsteroidal anti-inflammatory drugs using Medicaid data and from multiple other sources have produced similar results.⁴⁴ It is much more of a problem for descriptive studies.

DIAGNOSTIC TERMINOLOGY

Medicaid diagnosis data is coded using the International Classification of Disease Ninth Revision—Clinical Modification (ICD-9-CM) coding scheme. While this coding system is the most common scheme used for such purposes, its use can still be problematic.

First, there are often many different ICD-9-CM codes compatible with the same disease process. For example, upper gastrointestinal bleeding from a duodenal ulcer could be coded as upper gastrointestinal bleeding not otherwise specified (ICD-9-CM code 578.9), hematemesis (ICD-9-CM code 578.0), melena (ICD-9-CM code 578.1), acute duodenal ulcer with bleeding (ICD-9-CM code 532.0), peptic ulcer with bleeding (ICD-9-CM code 533.0), etc. Thus, it is necessary to combine many different ICD-9-CM codes into a single diagnostic code.

Second, there is no incentive for a provider to code specifically, e.g., duodenal ulcer with bleeding rather than peptic ulcer with bleeding or upper gastrointestinal bleeding not otherwise specified. Thus, one must be careful in interpreting such distinctions and, in general, researchers using claims data must be "lumpers" rather than "splitters."

Third, ICD-9-CM codes do not always quite fit the clinical syndrome of interest. One must be prepared to use multiple different aggregations of codes as different definitions of the same disease. For example, in a study of zomepirac and hypersensitivity reactions⁴⁵ the investigators used

six different definitions of hypersensitivity reactions:

1. bronchospasm and laryngospasm,
2. shock, other than septic and cardiogenic shock,
3. allergy unspecified,
4. allergic skin reactions (e.g., urticaria),
5. adverse effects of medicinal and biological substances, and
6. the subset of each of the first five codes which was felt to best approximate the syndrome of anaphylaxis without including too much non-anaphylactic disease.

Fourth, the ICD-9-CM coding scheme can sometimes be too inclusive. For example, the code for erythema multiforme (ICD-9-CM code 695.1) includes erythema multiforme major, erythema multiforme minor, Stevens–Johnson Syndrome, toxic epidermal necrolysis, staphylococcal scalded skin syndrome, etc., some of which may share common mechanisms but others of which probably do not. In this situation, one must use primary medical records to make the distinction.

Finally, outcomes that do not reliably result in medical encounters are subject to underascertainment. This applies most notably to quality of life endpoints.

CONFOUNDING VARIABLES

A confounding variable is a patient characteristic that is related to both the disease and the exposure under study in such a way that it could create a false association between the exposure and the disease or mask a real one. For example, if one were examining the relationship between alcohol consumption and lung cancer, a positive association might emerge. However, it has been observed that people who use alcohol tend to be smokers. If the relationship between alcohol and lung cancer were then re-evaluated accounting for this fact, the initial relationship might disappear. In this situation it would be possible to control for the confounding variable (smoking) by stratifying cases and controls according to smoking status and then testing to see whether the relationship between alcohol use and lung cancer persisted for

smokers and nonsmokers. This is discussed in more detail in Chapters 2 and 43.

In any epidemiologic study there will be a number of important potential confounding variables, including age, sex, geographic location, indications or contraindications for drug therapy, concomitant drug therapy, and underlying diagnoses. When information on the confounding variable is available, it can be controlled by using standard epidemiologic techniques: exclusion, matching, stratification, and/or mathematical modeling. Each of these, however, requires that one have information on that confounding variable. In Medicaid databases, those risk factors which are prescribed by physicians, such as drugs or procedures, or can be coded in ICD-9-CM format, such as diagnoses, can be controlled. However, as a billing database, Medicaid datasets lack information on some variables which can be very important as potential confounding variables. For example, smoking, environmental exposures, illicit drug use, alcohol use, occupation, family history, and use of over-the-counter drugs all are missing from Medicaid data.

Sometimes other variables can be used as proxies. As an example, one could use alcohol-related illnesses as a surrogate for alcoholism.¹³ These would include alcoholic cirrhosis, alcoholic hepatitis, delirium tremens, etc. The results would then need to be interpreted realizing that considerable misclassification remains, but that this is unidirectional misclassification. All those classified as alcoholics clearly would be, but many others would be alcohol abusers, as well. The control of some other factors, such as occupational exposures, is impossible.

Alternatively, information on confounders which is missing can be obtained from primary medical records. This approach has been used widely in recent studies.^{11–13,46} For example, in a study of drug induced liver disease, cases with a history of significant alcohol use in their medical record were excluded.¹² This approach will be completely successful only for variables for which the information is uniformly recorded in the medical record. In a study of subjects with liver disease, the use of alcohol will be almost always be recorded since alcohol is one of the most

common causes of liver disease. In contrast, smoking may not be reliably recorded in this study because this variable is not related to liver disease. However, in a study of lung cancer patients, smoking would almost always be recorded in the medical record.

This approach to control for confounding is limited if similar exclusions are not applied to controls. However, it is usually not possible to obtain the medical records for outpatient controls. The variable of interest is controlled for in the cases by exclusion. Since information on the variable in the controls is not available, it must be assumed that this variable is uncommon. If the variable is not uncommon, than other methods must be used to control for confounding.

Another approach is to assess the degree of potential confounding in the cases. This approach is applicable for risk factors identifiable from medical records and is based upon the fact that, in the absence of effect modification, a true confounder will have a different prevalence in exposed and unexposed cases.^{129,130} This approach was used in a study of nonsteroidal anti-inflammatory drugs and peptic ulcer disease and found that among cases of peptic ulcer disease, current NSAID users are not more likely to be smokers or to have a past history of peptic ulcer disease than are nonusers.^{20,30}

Other approaches include use of control groups as similar as possible to drug users by comparing drugs from the same therapeutic class^{29,47} or time-on:time-off drug.^{48,49} However, because confounders may be linked with choice of therapeutic agent (e.g., avoiding long acting benzodiazepines in frail patients) or with reasons for stopping drug (e.g., a prodromal symptom of the disease under study), this approach cannot rule out confounding.

Many of these techniques are used in a study. However, the potential for confounding by factors unavailable in a database sometimes may be so great that conducting a study is not advisable. For example, comparison of rates of suicide between users of different antidepressants could be substantially confounded by differences in severity of depression, a factor that is very difficult to reliably ascertain retrospectively.

DATA VALIDITY

The most important concern with using billing data is the validity of its data. An FDA-funded validation study of COMPASS[®] was completed in 1984, comparing Medicaid data from Michigan and Minnesota to its primary sources, i.e., data from hospitals, physicians, pharmacies, etc.⁵⁰ The results of this study indicated that the demographic and drug data appeared to be of extremely high quality. Within pre-established limits, year of birth agreed in 94% of sampled patients, and could not be determined from the medical records in another 2.5%; sex agreed in 95% of patients, and could not be determined from the medical records in another 4%; and the date of a pharmacy's dispensing of each drug agreed in 97% of sampled prescriptions. Regarding medical services, of those records that could be evaluated, 93% of the services in Medicaid data could be found in the provider records looking within 1 week of the Medicaid date; in 17% of those, the provider record included a previous or subsequent visit that was not in the Medicaid data. Diagnostic agreement to at least three digits of the ICD-9-CM code occurred in 41%, agreement within a broad diagnostic category in another 16% (i.e., same body system and/or type of illness), no diagnosis was present on the provider record in 12%, a single diagnosis in 3%, and there was no agreement in 28%. Clearly, this study suggested that diagnostic data validity needs to be considered in each individual study.

Several levels of validation need to be considered. The first is whether the computer diagnosis accurately reflects the medical records. This merely implies that there was not a coding error. The second level is whether the diagnosis made and recorded by the physician is correct. For example, a physician may diagnose a skin rash as erythema multiforme, when in fact it is some other skin disorder. Specific criteria need to be developed to validate diagnoses in this way. It is important to keep in mind that primary medical records may not include adequate detail to perform the second level of validation. However, validation efforts must keep each of these two different types of validation clearly in mind.

A series of formal validation studies have now been completed in COMPASS[®] using primary medical records. Samples of patients were selected with an inpatient diagnosis of erythema multiforme ($n = 168$), an inpatient diagnosis of neutropenia ($n = 198$), any diagnosis of hypersensitivity reactions ($n = 52$), and inpatient diagnosis of acute liver disease ($n = 414$), among others. Overall, a diagnosis compatible with the computerized claim diagnosis was present in approximately 95% of the medical records reviewed in detail.

Each of the specific studies, however, illustrated strengths and weaknesses of the data. In the neutropenia validation study, out of 198 medical charts that were reviewed, 192 contained information on white cell counts that confirmed the diagnosis of neutropenia (i.e., 97% agreement).¹⁰ The median total white cell count was 2700 with a minimum of 100 and a maximum of 8800. The few patients with a normal white cell count had a very low percentage of their white cells that were polymorphic leukocytes, i.e., their total neutrophil count was low, and, therefore, the diagnosis was correct. However, 13.5% of the cases had recurrent disease, and 9.9% had cyclic neutropenia. Thus, even when the computer diagnosis is highly accurate, other information from the medical record can be needed to characterize the cases accurately.

For the Stevens–Johnson Syndrome study, the primary medical records of 249 cases with an inpatient diagnosis of “erythema multiforme” (ICD-9-CM code 695.1) were sought from three states.⁵¹ Of these, 128 (51.4%) of medical records were received. Sufficient information to assess the accuracy of the diagnosis was found in 122 (95.3%). Of the 128 records that were reviewed, 121 (94.5%) had a skin diagnosis that could have been compatible with ICD-9-CM code 695.1. Of the 128 cases with a computer diagnosis of “erythema multiforme,” only 15.6% of patients had a discharge diagnosis of erythema multiforme major or minor. When these records were reviewed by a dermatologist, 14.8% had the diagnosis of erythema multiforme major confirmed and another 27.3% had erythema multiforme minor diagnosed. This analysis confirmed that this ICD-9-CM code includes many unrelated diag-

noses, and that the physicians’ discharge diagnosis is not always accurate.

For a study of drug induced acute liver disease, the medical records of 414 patients with a computer diagnosis of liver disease were reviewed.¹² One of the goals of the review was to exclude patients with identifiable causes of liver disease besides drugs. Of the records reviewed, 15.9% were alcoholics, 31.9% had acute hepatitis A or B, 13.5% were intravenous drug users, 8.2% had acute cholecystitis or choledocholithiasis, and 4.1% had received a transfusion within 6 months. No liver diagnosis was found in 10.6%, and 5.7% had chronic liver disease. Of the 169 cases of idiopathic acute liver disease identified, many were hospitalized for reasons besides liver disease, and had very mild disease. Thus, this study found that Medicaid billing data has high reliability and validity for the diagnosis of acute liver disease. However, primary medical records are essential for the study of drug induced hepatitis, to be able to exclude other causes of liver disease, and to obtain other information not included in the computer data.

Our experience suggests that, with few exceptions, medical records must be obtained in every study to confirm the diagnosis, characterize the severity of the disease, and to obtain information on potential confounding variables not found in the computer data. This is particularly problematic, since the New Jersey database never developed access to medical records, and COMPASS[®] has lost its access. In general, we have also found that outpatient records are difficult to obtain and often do not have the information necessary for the study. This suggests that studies using Medicaid data should be limited to diseases severe enough to result in hospitalization.

An exception when primary medical records may be less important is studies which focus on evaluating drug to drug relationships, or studies which can use drugs or procedures as markers of diagnoses. For example, studies that found an association between patients receiving a new prescription for levodopa containing medication and prior exposure to metoclopramide is probably valid, since drug exposure is accurate.⁵²

Table 19.3. Selected known associations reproduced using Medicaid data

Exposure	Disease
Alcoholism	UGI bleeding
Ampicillin	Skin rash
Antibiotics	Neutropenia
Antithyroid drugs	Neutropenia
Captopril	Neutropenia
Diabetes mellitus	Gallbladder disease
Diethylstilbestrol	Thromboembolism
Erythromycin	Liver disease
Female sex	Hypersensitivity reactions
Increased age	UGI bleeding
Infections	Neutropenia
Inflammatory bowel disease	Neutropenia
Iron	Hypersensitivity reactions
Male sex	UGI bleeding
Metoclopramide	Parkinson's disease
NSAIDs	Neutropenia
NSAIDs	Renal disease
NSAIDs	UGI bleeding
Obesity	Gallbladder disease
Oral contraceptives	Gallbladder disease
Oral contraceptives	Thromboembolism
Other bacterial infect	Erythema multiforme
Other viral infect	Erythema multiforme
Pancreatitis	Gallbladder disease
Penicillin	Hypersensitivity reactions
Phenothiazines	Liver disease
Phenothiazines	Parkinson's disease
Potassium chloride	UGI bleeding
Pregnancy	Gallbladder disease
Previous abdominal disease	UGI bleeding
Procainamide (sustained release)	Neutropenia
Diethylstilbestrol	Neutropenia
Splenomegaly	Neutropenia
Sulfa drugs	Thrombocytopenia
Sulfonamides	Liver disease
Inflammatory bowel disease	Neutropenia
Tetracyclines	Liver disease
Thiazides	Gallbladder disease
Thiazides	Gout
Ticrynafen	Liver disease
Trimethoprim/sulfamethoxazole	Erythema multiforme
Upper respiratory infection	Erythema multiforme
Urinary tract infection	Erythema multiforme
Valproic acid	Neutropenia
Zomepirac	Hypersensitivity reaction

Many known associations have been reproduced using Medicaid data, e.g., those presented in Table 19.3.

POTENTIAL BIASES

Other potential errors in classification, i.e., misclassification biases, may occur in a number of ways. Drug entries should be valid since drug identity determines the amount of pharmacy Medicaid reimbursement. This is carefully audited by the states, and was confirmed in the FDA's validation study. The problem of a drug not being recorded in a patient's file is minimized by the fact that a chronic drug may legally be dispensed for a maximum of 90 days; most drugs can only be dispensed for 30 days. A prescription refill requires a return visit to the pharmacy. For drugs that can be purchased over the counter, it is likely that some individuals classified as nonusers may actually be users. It is likely that patients from the socioeconomic group eligible for Medicaid are less likely to pay for over-the-counter medications if they are required chronically or in large doses. However, this cannot be guaranteed.

Conversely, it is virtually guaranteed that some individuals who are prescribed a drug do not take the drug as directed. There is no way other than examining refill frequency to ascertain patient compliance in Medicaid datasets, although one advantage of the billing data systems is that they provide data on prescriptions which have been dispensed by the pharmacist, rather than simply written for by the physician. Therefore, in studies comparing patients exposed to a drug to unexposed patients, the risk associated with a drug may be underestimated if noncompliance is a significant problem. Analogously, studies which compare the relative risk of two drugs may be biased if there is a difference in compliance. For example, the risk of peptic ulcer disease has been demonstrated to be higher in patients taking piroxicam (once per day) compared to ibuprofen (four times per day).³¹

ELIGIBILITY CHANGES

Finally, and importantly, if a patient loses eligibility for Medicaid benefits during the study,

it may appear as if the patient does not develop the outcome of interest when in fact the patient is no longer included in the database. This problem is minimized by applying eligibility requirements, as discussed in more detail above, or using the recipient files in the states where the information is accurate.

PARTICULAR APPLICATIONS

EXAMPLES OF MEDICAID BASED STUDIES

In this section, we describe several studies that illustrate the use of Medicaid data for pharmacoepidemiology studies.

Griffin *et al.* studied the association of NSAID use and fatal upper gastrointestinal hemorrhage in a case-control study of Medicaid patients over 60 years old.³¹ Cases were defined as patients who died from upper gastrointestinal bleeding and were confirmed by review of the medical record. Controls were matched for age, sex, race, calendar year, and nursing home status. Cases were 4.7 (95% confidence interval, 3.1–7.2) times more likely to have filled an NSAID prescription than the controls. In a second study that extended these observations, Griffin *et al.* published the largest case-control study of peptic ulcer disease.³¹ The study included 1415 hospitalized cases 65 years of age or older, and 7063 controls from Medicaid enrollees. The diagnosis of peptic ulcer disease was confirmed by review of the medical record. Cases were 4.1 (95% CI, 3.5–4.7) times more likely to be current users of NSAIDs than controls. A dose and duration effect was again demonstrated. Finally, in a subanalysis of these data, Piper *et al.* demonstrated that the relative risk of peptic ulcer disease associated with the use of corticosteroids was 2.0 (95% CI, 1.3–3.0).²¹ However, the risk in patients taking both NSAIDs and corticosteroids was 4.4 (95% CI, 2.0–9.7), while that in patients only taking corticosteroids was 1.1 (95% CI, 0.5–2.1).

In a second series of studies, Carson and colleagues assessed the incidence of hospitalization for acute symptomatic hepatitis and performed two case-control studies.^{13,46} The diagnosis of idiopathic liver disease was validated by review of

the medical record. The incidence of acute liver disease resulting in hospitalization was 2.2 per 100 000 per year. In a study of nonsteroidal anti-inflammatory drug induced acute liver disease, the odds ratio was 1.2 (95% CI, 0.5–2.8).¹⁴ These results suggest that liver disease from NSAIDs is an infrequent clinical problem. In a second study of antibiotic induced liver disease, several important associations were demonstrated.⁴⁶ For erythromycins the odds ratio was 5.2 (95% CI, 1.1–26.6), for sulfonamides the odds ratio was 11.4 (95% CI, 2.7–67.8), and for tetracycline the odds ratio was 5.2 (95% CI, 1.4–19.7). The results did not change substantively after adjustment for multiple potential confounders.

In another set of studies, in order to evaluate a series of *a priori* hypotheses about drug-induced neutropenia, a population based case-control study was undertaken using 1980–1985 COMPASS[®] data from six states. Cases hospitalized for neutropenia were compared to four times as many age-, sex-, and state-matched controls, randomly chosen from the same datasets. The frequency of exposures of interest in the 30 days before hospital admission in the cases was compared to the frequency in the identical time periods in the controls. The annual incidence rates for neutropenia and agranulocytosis ranged from 31.5 to 40.5 and 6.3 to 8.1 per million per year, respectively.¹⁰ When properly adjusted for demographic differences, the incidence rates observed were virtually identical to those found in the International Agranulocytosis and Aplastic Anemia Study (at 1/100 the cost). Increased risks (odds ratio (90% confidence interval)) of neutropenia were seen with viral infections (4.7 (90% CI, 1.7–12.9)), other hematological diseases (72.6 (90% CI, 45.2–116.7)), splenomegaly (51.4 (6.1–∞)), autoimmune diseases (9.5 (90% CI, 4.6–19.7)), sarcoidosis (27.4 (2.3–∞)), antibiotics other than cephalosporins (2.2 (90% CI, 1.5–3.2)), nonsteroidal anti-inflammatory drugs (4.2 (90% CI, 2.0–8.7)), antithyroid drugs (19.5 (90% CI, 1.1–∞)), and slow release procainamide (27.4 (90% CI, 2.3–∞)). Allopurinol, captopril, and valproic acid all had odds ratios greater than 2.5, but the results did not reach conventional levels of statistical significance. In a separate analysis, no association was demonstrated between cimetidine and neutropenia.⁵³

Avorn and colleagues have performed a series of studies examining prior exposure to drugs in patients prescribed drugs to treat Parkinson's disease for the first time. Patients started on levodopa containing drugs were three times more likely to have been exposed to metoclopramide (odds ratio = 3.09, 95% CI, 2.25–4.26) than nonusers of drugs to treat Parkinson's disease.⁵² In a similar study, neuroleptic exposure was 5.4 times more likely (95% CI, 4.8–6.1) in patients started on antiparkinsonian therapy than nonusers.⁵⁴ These studies have the advantage of primarily relying on the drug exposure information, reducing the importance of obtaining medical records to confirm the diagnosis.

Medicaid data have been used to evaluate drug utilization and cost related to drug therapy. For example, a study using NJ Medicaid data showed that only about half of patients were still using lipid lowering treatment five years after initiating treatment.⁵⁵ Another study showed that prior authorization requirement for the use of brand name nonsteroidal anti-inflammatory drugs reduced the expenditure for nonsteroidal anti-inflammatory drugs by 53%.⁵⁶ A third study found the excess cost of gastrointestinal disease associated with nonsteroidal anti-inflammatory drugs to be very expensive.⁵⁷

APPROPRIATE ROLE OF MEDICAID DATA IN PHARMACOEPIDEMIOLOGY

Medicaid data are especially useful for studies in several circumstances. Many postmarketing pharmacoepidemiology studies require very large sample sizes, since the outcome and/or drug exposure of interest is very uncommon. If a quick answer to a question is required for a study, then databases that use data already collected such as these are preferable. Similarly, if resources are restricted, then studies can generally be performed at significantly reduced cost compared to an *ad hoc* study requiring *de novo* data collection. Finally, billing databases are preferable when studies are especially prone to recall bias, such as a study of drug induced birth defects, or interviewer bias.

There are also several situations when Medicaid data should not be used for pharmacoepidemiology studies. First, Medicaid data cannot be used to study inpatient drug use. Second, Medicaid data cannot be used to study over-the-counter drug use. Third, Medicaid data should not be used to study diseases that will not reliably come to medical attention, such as minor nausea and mild skin rashes. Fourth, a Medicaid database should not be used if it lacks information about confounders important to that study. For example, it would not be appropriate to use Medicaid data to perform a study on lung cancer, since smoking history is unavailable. Fifth, Medicaid data should not be used if the ICD-9-CM coding system does not adequately describe the outcome of interest. An example would be a study of warfarin induced skin necrosis. There also are times when this coding system is too inclusive, so that related diseases are grouped together. An example would be a study of drug induced retroperitoneal fibrosis, since retroperitoneal fibrosis is grouped into the code for urethral obstruction. Sixth, Medicaid data may not be suitable to study long term effects of drugs if frequent eligibility changes result in many subjects losing benefits. Finally, Medicaid (or other automated databases) should not be used if crucial study variables need to be determined via patient contact, such as depression, hypertension, and lipid levels.

THE FUTURE

The three Medicaid databases commonly used for pharmacoepidemiology research are composed of Medicaid billing data from about three million patients. Other Medicaid databases are available as well, although they have not often been used for pharmacoepidemiology studies. Very large studies can be performed with these resources, relatively quickly and inexpensively. These comprehensive databases permit studies of both inpatient and outpatient diseases and permit calculation of incidence rates. Maternal–fetal linkage permits studies of birth defects.

The principal concern in using this type of database is the validity of the diagnosis data.

Recent experience documents that high quality studies can be performed when primary medical records are reviewed. The recent past has seen increasing use of these data for pharmacoepidemiology studies, with increasing awareness of those diagnoses that can be validly studied using these data versus those that cannot. The future is also likely to see the inclusion of data from more Medicaid programs, and more non-Medicaid programs. However, public concerns about protecting confidentiality have made it harder to obtain primary medical records. This is a problem which must be solved, if these resources are to continue to be most useful. However, overall, used with care, Medicaid databases appear to be potentially very useful, growing, and important sources of drug use/diagnosis experience data for the FDA, the pharmaceutical industry, and academic institutions.

REFERENCES

1. Waid, MO. *Brief Summaries of Medicare and Medicaid: Title XVIII and Title XIX of The Social Security Act as of June 25, 1998*. Health Care Financing Administration.
2. Lessler JT, Harris BSH. *Medicaid Data as a Source for Postmarketing Surveillance Information*. Research Triangle Park, NC: Research Triangle Institute, 1984.
3. Ray WA, Griffin MR. The use of Medicaid data for pharmacoepidemiology. *Am J Epidemiol* 1989; **129**: 837–49.
4. Carson JL, Strom BL, Morse ML. COMPASS[®] database: An epidemiologist's viewpoint. In: *Worldwide Drug Development and Scientific Data Management. The Impact of Emerging Technologies and Regulations. Proceedings of the Drug Information Association, Philadelphia, PA, 1984*. 97–9.
5. Strom BL, Carson JL, Morse ML, et al. The Computerized On-Line Medicaid Pharmaceutical Analysis and Surveillance System: a new resource for post-marketing drug surveillance. *Clin Pharmacol Ther* 1985; **38**: 359–64.
6. Morse ML, LeRoy AA, Strom BL. COMPASS[®]. A population-based postmarketing drug surveillance system. In: Inman WHW, ed., *Monitoring for Drug Safety*. Philadelphia, PA: Lippincott, 1986; 237–54.
7. Jones JK, Van de Carr SW, Zimmerman H, et al. Hepatotoxicity associated with phenothiazines. *Psychopharmacol Bull* 1983; **19**: 24–7.
8. Van de Carr SW, Kennedy DL, Rosa FW, Anello C, Jones JK. Relationship of oral contraceptive estrogen dose to age. *Am J Epidemiol* 1983; **117**: 153–9.
9. Strom BL, Tamragouri RN, Morse ML, Lasur EL. Oral contraceptives and other risk factors for gallbladder disease. *Clin Pharmacol Ther* 1986; **39**: 335–41.
10. Strom BL, Carson JL, Schinnar R, Sims E, Shaw M. Descriptive epidemiology of agranulocytosis. *Arch Intern Med* 1992; **152**: 1475–80.
11. Strom BL, Carson JL, Schinnar R, Sims E, Shaw M, Lundin FE. Nonsteroidal anti-inflammatory drugs and neutropenia. *Arch Intern Med* 1993; **153**: 2119–24.
12. Carson JL, Strom BL, Duff A, Gupta A, Das K. The feasibility of studying acute hepatitis using Medicaid data. *Clin Pharmacol Ther* 1992; **52**: 214–9.
13. Carson JL, Strom BL, Duff A, Gupta A, Das K. Safety of NSAIDs with respect to acute liver disease. *Arch Intern Med* 1993; **153**: 1331–6.
14. Carson JL, Strom BL, Schinnar R, Duff A, Sim E. The low risk of upper gastrointestinal bleeding in patients dispensed corticosteroids. *Am J Med* 1991; **91**: 223–8.
15. Ray WA. Unpublished data 1999.
16. Ruther M, Pagan BA, Rinkle V, Yanck J. *Program Statistics: Medicare and Medicaid Data Book, 1984*. Baltimore, MD: Health Care Financing Administration, 1986.
17. Mirvis DM, Chang CF, Hall CJ, Zaar GT, Applegate WB. TennCare—health system reform for Tennessee. *J Am Med Assoc* 1995; **274**: 1235–41.
18. Ray WA, Fought RL, Decker MD. Psychoactive drugs and the risk of injurious motor vehicle crashes in elderly drivers. *Am J Epidemiol* 1992; **136**: 873–83.
19. Ray WA, Gurwitz J, Decker MD, Kennedy DL. Medications and the safety of the older driver: is there a basis for concern? *Hum Factors* 1992; **34**: 33–47.
20. Shalat L, True LD, Fleming LE, Pace P. Kidney cancer in utility workers exposed to polychlorinated biphenyls (PCBs). *Br J Ind Med* 1989; **46**: 823–4.
21. Piper JM, Ray WA, Rosa FW. Pregnancy outcome following exposure to angiotensin-converting enzyme inhibitors. *Obstet Gynecol* 1992; **80**: 429–32.
22. Shorr RI, Ray WA, Daugherty JR, Griffin MR. Individual sulfonylureas and serious hypoglycemia in the elderly. *J Am Geriatr Soc* 1996; **44**: 751–5.
23. Ray WA, Federspiel CF, Schaffner W. Prescribing of chloramphenicol in ambulatory practice: An epidemiologic study among Tennessee Medicaid recipients. *Ann Intern Med* 1976; **84**: 266–70.
24. Ray WA, Federspiel CF, Schaffner W. Prescribing of tetracycline to children less than 8 years old: A two-year epidemiologic study among ambulatory

- Tennessee Medicaid recipients. *J Am Med Assoc* 1977; **237**: 2069–74.
25. Schaffner W, Ray WA, Federspiel CF. Surveillance of antibiotic prescribing in office practice. *Ann Intern Med* 1978; **89**: 796–9.
 26. Schaffner W, Ray WA, Federspiel CF, et al. Improved antibiotic prescribing in office practice. A controlled trial of three educational methods. *J Am Med Assoc* 1983; **250**: 1728–32.
 27. Ray WA, Fink R, Schaffner W, Federspiel CF. Improving antibiotic prescribing in outpatient practice: Nonassociation of outcome with prescriber characteristics and measures of receptivity. *Med Care* 1985; **23**: 1307–14.
 28. Ray WA, Schaffner W, Pates JA. Therapeutic choice in the treatment of hypertension. Initial treatment of newly diagnosed hypertension and secular trends in the prescribing of antihypertensive medication for medicaid patients. *Am J Med* 1986; **81** (suppl. 6C): 9–16.
 29. Ray WA, Griffin MR, Schaffner W, et al. Psychotropic drug use and the risk of hip fractures. *New Engl J Med* 1987; **316**: 363–9.
 30. Griffin MR, Ray WA, Schaffner W. Nonsteroidal anti-inflammatory drug use and death from peptic ulcer disease in the elderly. *Ann Intern Med* 1988; **109**: 359–63.
 31. Griffin MR, Piper JM, Daugherty JR, Snowden M, Ray W. Nonsteroidal anti-inflammatory drug use and increased risk for peptic ulcer disease in elderly persons. *Ann Intern Med* 1991; **114**: 257–63.
 32. Piper JM, Ray WA, Daugherty JR, Griffin MR. Corticosteroid use and peptic ulcer disease: Role of nonsteroidal anti-inflammatory drugs. *Ann Intern Med* 1991; **114**: 735–40.
 33. Griffin MR, Ray WA, Mortimer EA, Fenichel GM, Schaffner W. Risk of seizures after measles–mumps–rubella immunization. *Pediatrics* 1991; **88**: 881–5.
 34. Ray WA, Griffin MR, Baugh DK. Mortality following hip fracture before and after implementation of the prospective payment system. *Arch Intern Med* 1990; **150**: 2109–14.
 35. Shorr RI, Ray WA, Daugherty JR, Griffin MR. Individual sulfonylureas and serious hypoglycemia in older people. *J Am Geriatr Soc* 1996; **44**: 751–5.
 36. Shorr RI, Ray WA, Daugherty JR, Griffin MR. Antihypertensives and the risk of serious hypoglycemia in older persons using insulin or sulfonylureas. *J Am Med Assoc* 1997; **278**: 40–3.
 37. Soumerai SB, Ross-Degan D, Gortmaker S, Avorn J. Withdrawing payment for nonscientific drug therapy. Intended and unexpected effects of a large-scale natural experiment. *J Am Med Assoc* 1990; **263**: 831–9.
 38. Soumerai SB, Ross-Degan D, Avorn J, McLaughlin TJ, Choodnovskiy I. Effects of Medicaid drug-payment limits on admission to hospitals and nursing homes. *New Engl J Med* 1991; **325**: 1072–7.
 39. Avorn J, Everitt DE, Weiss S. Increased antidepressant use in patients prescribed beta-blockers. *J Am Med Assoc* 1986; **225**: 357–60.
 40. Gurwitz JK, Bohn RL, Glynn RJ, Monane M, Morgun H, Avorn J. Antihypertensive drug therapy and the initiation of treatment for diabetes mellitus. *Ann Intern Med* 1993; **118**: 273–8.
 41. Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.
 42. Thapa PB, Whitlock JA, Brockman KG, Gideon P, Mitchel EF, Roberson P, Pais R, Ray WA. Prenatal exposure to metronidazole and risk of childhood cancer. *Cancer* 1998; **83**: 1461–8.
 43. Brown NJ, Ray WA, Snowden M, Griffin MR. Black Americans have an increased rate of angiotensin converting enzyme inhibitor-associated angioedema. *Clin Pharmacol Ther* 1996; **60**: 8–13.
 44. Henry D, Lim LLY, Garica Rodriguez LA, Perez Gutthann S, Carson JL, et al. Variability in risk of gastrointestinal complications with individual nonsteroidal anti-inflammatory drugs. Results of a collaborative meta-analysis. *Br Med J* 1996; **312**: 1563–6.
 45. Strom BL, Carson JL, Morse ML, et al. Hypersensitivity reactions associated with zomepirac sodium and other non-steroidal anti-inflammatory drugs. *Arthritis Rheum* 1987; **30**: 1142–8.
 46. Carson JL, Strom BL, Duff A, Gupta A, Das K. Acute hepatitis associated with erythromycins, sulfonamides, and tetracyclines. *Ann Intern Med* 1993; **119**: 576–83.
 47. Ray WA, Griffin MR, Downey W. Benzodiazepines of long and short elimination half-life and the risk of hip fracture. *J Am Med Assoc* 1989; **262**: 3303–7.
 48. Ray WA, Fought RL, Decker MD. Psychoactive drugs and the risk of injurious motor vehicle crashes in elderly drivers. *Am J Epidemiol* 1992; **136**: 873–83.
 49. Wyndaele J-J, Maes D. Clean intermittent self-catheterization: A 12-year followup. *J Urol* 1990; **143**: 906–8.
 50. Lessler JT, Harris BSH. *Medicaid Data as a Source for Ppost Marketing Surveillance Information*, final report. Research Triangle Park, NC: Research Triangle Institute, 1984.
 51. Strom BL, Carson JL, Halpern AC, Schinnar R, Snyder ES, Shaw M, Tilson HH, Joseph M, Dai WS, Chen D, Stern RS, Bergman U, Stolley PD. A population-based study of Stevens–Johnson Syndrome: Incidence and antecedent drug exposures. *Arch Dermatol* 1991; **127**: 831–8.
 52. Avorn J, Gurwitz JH, Bohn RL, Mogun H, Monane M, Walker AM. Increased incidence of levodopa therapy following metoclopramide use. *J Am Med Assoc* 1995; **274**: 1780–2.

53. Strom BL, Carson JL, Schinnar R, Shaw M. Is cimetidine associated with neutropenia? *Am J Med* 1995; **99**: 282–90.
54. Avorn J, Bohn RL, Mogun H, Gurwitz JH, Monane M, Everitt D, Walker A. Neuroleptic drug exposure and treatment of parkinsonism in the elderly: a case–control study. *Am J Med* 1995; **99**: 48–54.
55. Avorn J, Monette J, Lacour A, Bohn RL, Monane M, Mogun H, *et al.* Persistence of use of lipid-lowering medications: a cross-national study. *J Am Med Assoc* 1998; **279**: 1458–62.
56. Smalley WE, Griffin MR, Fought RL, Sullivan L, Ray WA. Effect of a prior-authorization requirement on the use of nonsteroidal antiinflammatory drugs by Medicaid patients. *New Engl J Med* 1995; **332**: 1612–7.
57. Smalley WE, Griffin MR, Fought RL, Ray WA. Excess costs from gastrointestinal disease associated with nonsteroidal anti-inflammatory drugs. *J Gen Intern Med* 1996; **11**: 461–9.

20

Health Databases in Saskatchewan

WINANNE DOWNEY, PATRICIA BECK, MARY McNUTT, MARYROSE STANG,
WILLIAM OSEI AND JIM NICHOL

Research Services, Population Health Branch, Saskatchewan Health, Regina, Saskatchewan, Canada

INTRODUCTION

Saskatchewan is one of ten provinces and three territories in Canada. It is located in western Canada and has a stable population of about one million people, or 3.4% of the total population of Canada.

Saskatchewan has a publicly funded health system. Within this system, Saskatchewan Health, a provincial government department, and 32 district health boards (local health authorities) provide health services to the citizens of Saskatchewan. With funding from Saskatchewan Health, the district boards plan and deliver most services to people within their geographic jurisdictions based on the needs of their residents. Saskatchewan Health coordinates province-wide programs such as the Prescription Drug Plan and physician services.

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, Saskatchewan Health has been accumulating a very large amount of health care information in computerized databases over a number of years. These

databases have been recognized as a resource for drug utilization review and pharmacoepidemiology.¹⁻⁸ Publications of studies based on the data reflect the value of the databases for this research.⁹⁻⁸¹

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 information. These and some of the other health databases are described in more detail below.

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 resident. 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 (the "covered population"). As of 30 June, 1998, more than one million people were eligible for health benefits (Table 20.1).⁸² More than 2.2 million individuals have been registered for coverage since 1968. Excluded from eligibility, and therefore from the population registry, are people whose health care is fully funded federally. This category, which includes members of the Royal Canadian Mounted Police, members of the Canadian Forces, and inmates of federal penitentiaries, accounts for less than 1% of the total population. Consequently, the population registry (together with specific health services information) enables study of all segments of the population including children, women of childbearing age, and the elderly.

The population registry captures demographic and coverage data on every member of the eligible population (Table 20.2). It is updated daily for name or address changes, births, deaths, new residents, departing residents, and those qualifying for social services supplementary health coverage. The registry is verified and updated through continuous, routine checks by field staff and by a variety of mechanisms which signal changes in eligibility, identification, or demographic data. Transactions for an insured health service are checked against the population registry for elig-

Table 20.1. Eligible population by age and sex: 30 June, 1998

Age (years)	Females	Males	Total
Under 1	6 061	6 151	12 212
1-9	65 828	69 066	134 894
10-19	78 721	82 985	161 706
20-29	68 075	71 011	139 086
30-39	76 811	76 802	153 613
40-49	71 430	74 052	145 482
50-59	47 921	48 728	96 649
60-69	39 971	38 569	78 540
70-79	37 142	30 157	67 299
80-89	22 176	13 604	35 780
90+	4 614	2 058	6 672
Total	518 750	513 183	1 031 933

Table 20.2. Information contained in the population registry

Name
Health services number
Sex
Marital status
Date of birth
Date of death (if applicable)
Mailing address
Location code
Indicator for registered Indian status
Indicator for current recipients of Saskatchewan Assistance Plan (welfare)
Dates of coverage initiation and termination

ibility of the claimant and for accuracy of identification and demographic information; inconsistencies are manually checked.

PRESCRIPTION DRUG DATA

All Saskatchewan residents are eligible for benefits under the Prescription Drug Plan (the "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. (These ineligible individuals can be excluded from studies.)

Drugs covered by the Drug Plan are listed in the Saskatchewan Formulary; nonformulary drugs generally are not covered, although there are exceptions. All drugs listed in the Saskatchewan Formulary are intended for outpatients. (This database also includes prescriptions dispensed to most residents of long term care facilities, however.) The formulary is updated semi-annually; as of July 1998, it included 2875 drug products. Listed drugs are under continuous review; new drug products are evaluated as applications are made and are added if they meet the standards of professional expert committees that aid the provincial government. Drugs must be prescribed by a licensed practitioner in order to be covered.

The Drug Plan began providing benefits to residents on 1 September, 1975. When the program began, consumers paid a portion of the professional fee, and Saskatchewan Health paid the pharmacy for the remainder of the prescription

cost; those individuals who could not afford to pay even a portion of the prescription fee were entitled to benefits without any payment. Currently, a family based deductible system is in place.

Since January 1989, pharmacies have submitted claims to Saskatchewan Health through point-of-service (POS) computer terminals, connected via telecommunication lines to a central database. All prescription data are collected (including those whose costs do not exceed the deductible) and data are collected on an individual patient basis.

There are a number of validation checks made of the data. The computerized, claim processing system checks each transaction for identification and demographic accuracy, claimant eligibility, and drug coverage under the Drug Plan. The immediate visual feedback incorporated in the POS computer-telecommunication processing system provides an additional verification check at the pharmacy level during data entry. Pharmacy audits are conducted on a regular basis. In addition, drug utilization review and program evaluations using the drug data are ongoing within the Drug Plan; to some extent, this data usage affords validation.

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.) It includes patient, drug, prescriber, pharmacy, and cost information (Table 20.3). The drug classification schemes used enable analysis of drugs at a brand, generic, or class level; customized programs facilitate categorization into various therapeutic and/or pharmacological groups. Because the patient is identified by the HSN, the drug data can be linked to the population registry to compile information on the age, sex, and residence of users of a product. The prescriber number can be linked with a physician registry for additional information, such as location of practice and prescriber specialty. The pharmacy number enables classification by location, type of pharmacy (chain or independent), and prescription volume. The drug database does not include information on most nonformulary prescription drug use, most over-the-counter drug use, use of professional samples, and in-hospital medications. In addition, it does not include

Table 20.3. Information contained in the prescription drug database

<i>Patient information</i>
Health services number
Sex
Year of birth; age
Designation of special status (e.g., welfare recipient, palliative care registrant, long-term care home resident)
<i>Drug information</i>
Pharmacologic-therapeutic classification ^a
Drug identification number (DIN) ^b
Active ingredient number of drug (AIN) ^b
Generic and brand names
Strength and dosage form
Manufacturer of drug
Date dispensed
Quantity dispensed
"No substitution" indicator, if applicable
<i>Prescriber information</i>
Prescriber identification number
<i>Dispensing pharmacy information</i>
Pharmacy identification number
<i>Cost information</i>
Unit cost of drug materials
Dispensing fee
Markup
Consumer share of total cost
Drug Plan share of total cost
Total cost

^aThe American Hospital Formulary Service classification system is used.

^bAssigned by the Health Protection Branch of Health Canada.

information about the dosage regimen prescribed, the reason the drug was prescribed, or patient compliance.

In the fiscal year 1997-98, more than 6.2 million prescription claims were processed by the Drug Plan (Table 20.4). Sixty-six percent of those eligible received benefits; individuals 65 years and older made up 15.5% of the eligible population and received over 45% of all prescriptions processed under the Drug Plan.

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 and day surgery case, thus creating a

Table 20.4. Prescriptions covered by the drug plan by pharmacologic–therapeutic classification: 1997–98

Pharmacologic–therapeutic classification ^a	Number of prescriptions ^b
8:00 Anti-infectives	745 730
10:00 Antineoplastic drugs	1 308
12:00 Autonomic drugs	246 781
20:00 Blood formation and coagulation	96 618
24:00 Cardiovascular drugs	1 441 984
28:00 Central nervous system drugs	1 238 508
36:00 Diagnostic agents	68 974
40:00 Electrolytic, caloric and water balance	380 588
48:00 Cough preparations	506
52:00 Eye, ear, nose and throat preparations	243 339
56:00 Gastrointestinal drugs	351 746
60:00 Gold compounds	781
64:00 Metal antagonists	735
68:00 Hormones and substitutes	933 133
76:00 Oxytocics	194
84:00 Skin and mucous membrane preparations	236 470
86:00 Spasmolytics	40 111
88:00 Vitamins	45 508
92:00 Unclassified	188 153
Total	6 261 167

^aThe American Hospital Formulary Service classification system is used.

^bRefers to formulary and exception drug status prescriptions.

Provided by: Drug Plan and Extended Benefits Branch, Saskatchewan Health. Regina, Saskatchewan.

central databank of diagnostic and other health information on a defined population. (In Saskatchewan, a separation is defined as the discharge, transfer, or death of an inpatient.)

Five large hospitals provide a full range of hospital services to the populace in their immediate areas, serve as the tertiary referral centers for the province, and play a major role in research and education. There is also one rehabilitation hospital in Saskatchewan that offers physical and occupational rehabilitation. Seven regional hospitals serve as referral centers for the people from outside their immediate vicinity, provide basic hospital care to residents in their immediate locale, and assume a limited role in education. The largest number of general hospitals are classified as district or community hospitals that serve the local populace and offer a more limited range of services than do the base or regional hospitals.

Data are collected from all hospitals in the province. Included in the database are all acute care inpatient separations, day surgeries, long term care separations on patients who occupy a bed in a

general hospital, inpatient psychiatric separations on patients treated in general hospitals, active rehabilitation, and out-of-province hospital separations involving a member of the covered population.

All health service transactions are cross-checked with the population registry for patient eligibility and for identification and demographic accuracy. There are computer programs designed to detect illogical entries. Beginning in 1984–85, diagnostic data for over one-half of the admission/separations were processed by a national body, currently known as the Canadian Institute for Health Information (CIHI); as of 1998–99, all separations are being processed by CIHI. CIHI assigns a case mix group (CMG) and resource intensity weight (RIW) to each hospital separation, based on the discharge diagnoses and procedures performed in hospital. The CMG and RIW values enable estimation of costs associated with a given hospitalization.

The hospital separation database includes patient, diagnostic, treatment, and other information

Table 20.5. Information contained in the hospital separation database

Patient Information	Health Services Number Sex Month and year of birth
Diagnostic and treatment information	Up to three discharge diagnoses (4-digit ICD-9) Up to three procedures (4-digit CCP) Accident code (ICD-9 external cause code)
Other	Admission date Discharge date Level of care codes Length of stay Admission and separation types Case mix group Resource intensity weight Attending physician Attending surgeon (if applicable) Hospital identification number

(Table 20.5). Standard diagnostic and procedure classification systems are used. The diagnostic classification system is the World Health Organization International Classification of Diseases, Ninth Revision (ICD-9).⁸³ The procedure classification system is the Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures (CCP).⁸⁴ Both diagnoses and procedures are coded to four digits. Diagnostic coding for the majority of hospital discharges is undertaken at the hospital level, usually by health records administrators. Limited routine validation of the accuracy of this coding procedure is undertaken centrally. The data are accessible electronically from 1970 to the present. Data from hospitals are available approximately six months from date of discharge.

Inpatient medical records are stored in the hospitals. Hospitals must retain a patient's file for a minimum period of ten years from the date of last discharge or until age 19 if the patient is a minor, whichever period is longer, and must retain the record in its original form for a minimum period of six complete years.

In 1997–98, there were 159 800 separations for adults and children, 78 900 day surgery cases, and about 13 000 newborn records (including those born out-of-province). Table 20.6 provides a breakdown of the inpatient separations by ICD-9 diagnostic chapter.

PHYSICIAN SERVICES DATA

In Saskatchewan, most physician services are an insured benefit. All members of the covered population are eligible to receive benefits. Most physicians are reimbursed on a fee-for-service basis. 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). As of March 1998, 40% of the province's approximately 1130 active physicians were specialists; 60% were family practitioners. A beneficiary may seek services from any physician desired, while physicians retain the ability to accept or decline any patient.

Physicians submit claims to Saskatchewan Health for payment for services provided. Claims received from physicians contain patient identifiers, details to support payment, and the primary diagnosis of the patient's condition. Submitted claims are subjected to a series of computer checks to determine the validity of the claim data. Transactions are cross-checked with the population registry for patient eligibility and identification. Other internal computer checks are in place to check accuracy and to detect illogical entries. Claims rejected during the computer processing are reviewed manually.

Considerable patient, physician, diagnostic, and service information is available by using the claims file and the physician registry file (Table 20.7). Diagnoses are recorded using three-digit ICD-9 codes. (ICD-8 coding was used from 1971 to 1979.) Procedures are coded using fee-for-service codes from a payment schedule. Fee-for-service codes are established through consultation between Saskatchewan Health and the provincial 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,

Table 20.6. Hospitalizations by ICD-9 diagnostic chapter: 1997–98

ICD-9 Chapter	ICD-9 Codes	Separations		Patient days	
		Number	Per 1000 covered	Number	Per 1000 covered
I. Infectious and parasitic diseases	001–139	2 521	2.5	14 503	14
II. Neoplasms	140–239	8 727	8.6	76 457	75
III. Endocrine, nutritional and metabolic diseases, and immunity disorders	240–279	3 755	3.7	27 563	27
IV. Diseases of the blood and blood-forming organs	280–289	1 259	1.2	7 854	8
V. Mental disorders	290–319	3 610	3.5	28 764	28
VI. Diseases of the nervous system and sense organs	320–389	3 864	3.8	28 649	28
VII. Diseases of the circulatory system	390–459	20 899	20.5	147 966	145
VIII. Diseases of the respiratory system	460–519	17 560	17.2	91 587	90
IX. Diseases of the digestive system	520–579	18 818	18.4	84 180	83
X. Diseases of the genitourinary system	580–629	9 878	9.7	41 163	40
XI. Complications of pregnancy, childbirth, and the puerperium	630–676	17 019	16.7	51 883	51
XII. Diseases of the skin and subcutaneous tissue	680–709	2 077	2.0	13 866	14
XIII. Diseases of the musculoskeletal system and connective tissue	710–739	7 326	7.2	44 567	44
XIV. Congenital anomalies	740–759	990	1.0	5 491	5
XV. Conditions originating in the perinatal period	760–779	3 480	3.4	24 059	24
XVI. Symptoms, signs, and ill defined conditions	780–799	9 754	9.6	36 079	35
XVII. Injury and poisoning	800–999	12 953	12.7	71 700	70
Supplementary classification of factors influencing health status and contacts with health services	V01–V80	15 358	15.1	84 931	83
Total		159 848	156.7	881 262	864

Provided by: Corporate Information and Technology Branch, Saskatchewan Health. Regina, Saskatchewan.

health outcome studies should not be performed with unvalidated physician services data alone. When used in conjunction with the other databases, the outpatient physician services data can, however, play a useful role in data linkage projects. Claims data are accessible electronically from 1971 to the present. Data are available on most services within 90 days of the date the service was provided.

In 1997–98, more than 10 million services were provided by physicians. Table 20.8 shows the distribution of these services by ICD-9 diagnostic chapter.

CANCER SERVICES DATA

The Saskatchewan Cancer Program encompasses prevention, early detection, diagnosis, treatment, and followup of patients with malignant or premalignant 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. Approximately 95% of cancer notifications are from specialist referral or a pathology report, 5% are from death registrations or autopsy (singly or combined), and a small number are through physician claims. *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 through the appropriate provincial cancer clinic. All cases of invasive cancer are maintained on a followup program for a minimum of 10 years. The rate of loss to followup is approximately 3%.

The population based registry was established in 1932. Complete computerized data for all cancer sites are available since 1967; pre-1967 data are being computerized retroactively. Breast cancer data are now computerized from 1960 to the present. The cancer registry contains identification, case, death, and review information (Table 20.9) and can be related to radiotherapy and in-clinic chemotherapy treatment data. The registry increases by about 7000 cases per year (including nonmelanoma skin cancer and *in situ* cancers). Both confirmed and suspected cases of cancer are registered. The database includes approximately 160 000 individual patients and over 200 000 case records (the difference indicates the number of patients with cancer of two or more primary sites). Data are available within six months from the date of diagnosis of the cancer.

In 1997, the overall cancer incidence rates were 455.8 per 100 000 for males and 392.4 for females. Incidence rates by cancer site are shown in Table 20.10.

Table 20.7. Information available on physicians and physician services

Patient information	Health services number
	Age
	Sex
	Location of residence
Physician information	Indicator for registered Indian status
	Physician specialty
	Referring physician
	Clinic
	Age
	Sex
	Place and year of graduation
	Practice type (e.g., solo, association, rural, urban)
Diagnostic and service information	Date of service
	Service code
	Type of service
	Primary diagnosis (3-digit ICD-9 code)
	Location of service (e.g., office, inpatient, outpatient, home, other)
	Billing information (amount paid, date of payment)

VITAL STATISTICS

All birth, death, stillbirth, and marriage data are collected by Saskatchewan Health. Electronic data are readily accessible from 1974 to the present. Cause of death is recorded by a physician or coroner and is coded as received on a death registration form. If updated information such as an autopsy diagnosis is received, it takes priority

Table 20.8. Physician services by ICD-9 diagnostic chapter: 1997–98

ICD-9 chapter	ICD-9	Services		Patients	
		Number	Per 1000 covered	Number	Per 1000 covered
I. Infectious and parasitic diseases	001–139	309 996	303.8	125 287	122.8
II. Neoplasms	140–239	205 712	201.6	49 597	48.6
III. Endocrine, nutritional and metabolic diseases, and immunity disorders	240–279	306 458	300.3	91 395	89.6
IV. Diseases of the blood and blood-forming organs	280–289	68 834	67.5	22 967	22.5
V. Mental disorders	290–319	530 387	519.8	104 316	102.2
VI. Diseases of the nervous system and sense organs	320–389	856 990	839.9	241 636	236.8
VII. Diseases of the circulatory system	390–459	860 019	842.9	142 925	140.1
VIII. Diseases of the respiratory system	460–519	1 254 686	1 229.7	365 836	358.5
IX. Diseases of the digestive system	520–579	325 996	319.5	115 948	113.6
X. Diseases of the genitourinary system	580–629	530 742	520.2	156 805	153.7
XI. Complications of pregnancy, childbirth, and the puerperium	630–676	78 584	77.0	16 861	16.5
XII. Diseases of the skin and subcutaneous tissue	680–709	438 880	430.1	171 570	168.1
XIII. Diseases of the musculoskeletal system and connective tissue	710–739	539 255	528.5	197 567	193.6
XIV. Congenital anomalies	740–759	23 822	23.3	6 993	6.9
XV. Conditions originating in the perinatal period	760–779	12 156	11.9	4 094	4.0
XVI. Symptoms, signs, and ill defined conditions	780–799	496 492	486.6	179 707	176.1
XVII. Injury and poisoning	800–999	738 055	723.3	236 262	231.5
Supplementary classification of factors influencing health status and contacts with health services	V01–V80	1 103 090	1 081.1	301 115	295.1
All other diagnoses		1 367 315	1 340.0	327 264	320.7
Total		10 047 469	9 847.1	855 486 ^a	838.4

^a Does not equal column total; some patients are counted in more than one chapter.

Provided by: Medical Services and Health Registration Branch, Saskatchewan Health, Regina, Saskatchewan.

Table 20.9. Basic information contained in the cancer registry

Patient information	Health services number
	Name
	Sex
	Date and place of birth
	Address
	Marital status
Case information	Indicator for rural or urban residence
	Registration information (including tentative diagnosis, height, and weight)
	Final diagnostic information (ICD-O, behaviour, grade)
	Staging information
	Metastases at diagnosis
	Dates of diagnoses
	Method of diagnosis
	Pathology report and hospital record numbers
	Summary treatment at diagnosis (within four months)
	Performance status (functional level)
	Disease status
	Followup type
	Death information
Primary and secondary cause of death	
Autopsy status	
Disease status at death	
Review information	Review date
	Physician identifier
	Type of review
	Recurrence information
	Metastases information
	Treatment information
	Performance status
	Disease status
Height and weight	

Table 20.10. Incidence rates per 100 000 by cancer site: 1997

Cancer site	Incidence rate per 100 000	
	Males	Females
Lip	8.5	1.4
Oral cavity	2.2	2.3
Head and neck	4.1	2.5
Esophagus	6.5	3.1
Stomach	11.8	6.6
Colon	38.7	35.0
Rectum	23.7	14.0
Liver, gall bladder and biliary tract	6.5	4.7
Pancreas	8.1	9.7
Digestive tract	1.8	1.8
Larynx	6.3	1.0
Trachea, bronchus, lung, and respiratory system	72.8	39.9
Bone and connective tissue	3.2	3.1
Malignant melanoma of skin	10.5	11.9
Breast ^a	—	122.4
Cervix: invasive	—	12.3
Uterus	—	20.8
Ovary	—	16.7
Female genital organs	—	4.3
Prostate	120.8	—
Male genital organs	6.1	—
Kidney	13.0	7.6
Bladder	30.2	11.7
Other urinary tract	3.0	1.0
Brain and central nervous system	7.5	4.5
Thyroid and other endocrine glands	3.2	6.4
Lymphoma	21.7	16.3
Hodgkin's disease	1.8	3.7
Multiple myeloma	5.1	4.5
Leukemia	20.5	9.5
Primary site unknown	14.8	10.5
Other primary sites	2.6	3.1
Total	455.8	392.4

^a There were fewer than five cases of breast cancer among males in 1997. Provided by: Saskatchewan Cancer Agency, Regina, Saskatchewan.

over the registration form. (Autopsy reports are submitted voluntarily, but most are received.) The underlying cause of death is recorded electronically and is defined as the disease or injury that initiated the sequence of events that led to death. Four-digit ICD-9 coding is used.

Live births are registered by the family. Live birth registrations record obstetrical information

and gestational age. 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, will be found in the hospital services database because most births (over 98%) occur in

a hospital. Stillbirth registrations include a "medical certificate" section, which is completed and signed by a physician or coroner.

Although Vital Statistics data prior to 1992 do not include the HSN, those records can be searched by name or by the Vital Statistics registration number for linkage with other databases.

OTHER SASKATCHEWAN HEALTH INFORMATION

Other data are available that can be either linked with the HSN or manually reviewed to provide additional information on study populations. The following is a brief description of these databases.

Supportive Care Services

Data on institutional long term care and home care services delivered by district health boards are available in electronic form since April 1987, but are most accessible from April 1989.

Mental Health Services

Comprehensive mental health services including inpatient psychiatric care and community based outpatient services are provided throughout the province. All members of the covered population are eligible for benefits. Data on insured services are available in electronic form since 1967, but are most accessible from 1979 to 1996 for community based outpatient services and from 1979 to present for inpatient psychiatric care. Outpatient data are not available for 1997 or 1998 and may be limited in 1999. The mental health services database is currently undergoing a platform conversion which, when fully operational, should result in more timely data capture and improved data quality.

Alcohol and Drug Abuse Treatment Data

Some information is available on alcohol and drug abuse services provided by provincially funded programs. Computerized data are available since 1984–85.

Laboratory Records

A large volume of microbiologic and biochemical testing is centralized at the Saskatchewan Health Provincial Laboratory. Since October 1996, all tests performed centrally have been captured electronically. For some tests, this is the only laboratory in the province where the tests can be done, so the data are complete. These tests include neonatal screening, prenatal screening, chlamydia, pertussis, several viruses, DNA markers, and some hormones. The data are linkable with the other health databases and quality control checks are under way.

Medical Records

Hospital record abstraction has been used for a number of studies to collect additional information to complement or validate information derived from the administrative databases.^{17–21, 30–39, 44–51, 56–64, 69, 71, 79} Medical records in hospitals are accessible upon approval from individual district health boards and affiliated facilities. To date, all those approached have permitted use of their records. Personnel under contract with Saskatchewan Health access records for the purpose of abstraction. Positive identifiers are available to the abstractor but are removed before information leaves the facility. The final records are either unidentified or coded with a pseudo-identifier. Record retrieval rates have been excellent and typically exceed 95%.

Primary health records held by physicians have been accessed for research, but experience is limited. The process has been varied and has included both Saskatchewan Health abstractors visiting physician clinics and physician self-reporting. To date, physician participation and record retrieval rates have not approached the levels achieved with institution held records. As an example, in one recent study on antidepressant use, 58% of physicians approached agreed to participate; among participating physicians, approximately 88% of the required charts were available.⁷⁹ Lower rates of availability reflect the more dynamic nature of individual physician practice.

STRENGTHS

SUBJECT IDENTIFICATION

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 and hence can be used to link data from any of the computerized databases electronically.

POPULATION BASED DATA

Saskatchewan is a geopolitical entity composed of over 1 million people; over 2.2 million individuals have been covered in the registry over its tenure. The registry is updated regularly; therefore, it is very useful for providing current, valid denominator data.

ELECTRONICALLY LINKABLE DATA

Data housed within Saskatchewan Health are electronically linkable. This means that information can be compiled across the databases and over time. It also means that data can be merged and sorted on the basis of age, sex, geographic location, diagnosis, and a variety of other parameters.

CROSS-SECTIONAL AND LONGITUDINAL STUDIES

It is possible to carry out both cross-sectional and longitudinal outcome studies by linking data from two or more databases. Given the long tenure of the databases, it is possible to compile information about prior drug use and previous disease experience for study populations.

OUTPATIENT PRESCRIPTION DRUG DATA

The Saskatchewan Formulary is extensive and lists over 2800 drug products (as of 1998). The database captures the majority of prescription drug use in the province.

DIAGNOSTIC CODING WITH THE INTERNATIONAL CLASSIFICATION OF DISEASES

Both the hospital separation and the physician services data use the ICD-9 coding system for diagnoses. The hospital data include up to three discharge diagnoses coded at the four-digit level. The physician services data include one three-digit code per visit. The advantage of using the standard international coding system is that other agencies and other researchers can compare information from Saskatchewan with that from other jurisdictions. Adoption of the ICD-10 coding is currently planned for 1 April, 2001.

MEDICAL RECORD ACCESSIBILITY

Medical records in hospitals are accessible upon approval from individual district health boards and affiliated facilities. Cooperation by individual health districts and affiliated facilities to allow access to records has been excellent. Availability of records in the institutions has ranged from 95 to 100%, depending on the age of the record.

Physician records may also be accessed for specific studies. Access to physician records and the availability of information from them, however, are more variable than from hospital records.

DATA VERIFIED AND VALIDATED

For pharmacoepidemiology studies, it is important that data validity and reliability be evaluated. Data integrity is the responsibility of the respective administrative programs within Saskatchewan Health. The various claims processing systems have built in audit and eligibility checks.

For research purposes, further validation is necessary. The process of linking the various databases for a specific study provides a validation or reliability check for certain variables. For example, using data pooled from several studies, there was 96% agreement between the death flag on the registry and the death indicator on the hospital file.³⁵

Validation, mostly by hospital chart review, has been built into a number of studies using

Saskatchewan data. A wide range of conditions has been validated in this way, including rheumatoid arthritis,^{36,37} hip fractures,^{18,19,21,44} gastrointestinal bleeding,^{17,60,62,69} asthma-related conditions,^{30,31,38,39,45-48,56-59} puerperal seizures,²⁰ liver injury,^{33,34} and acute renal failure.⁶³ The validity of the diagnostic data is generally very good, but is dependent upon the ability of a diagnostic code to properly represent the condition in question and therefore varies with the condition. For example, for hip fracture diagnoses the sensitivity was over 99%,¹⁸ whereas for gastrointestinal bleeding and/or perforation it was 77%¹⁷ based on the particular case definition used in the study. Therefore, validity should be quantified for each condition studied.

Rawson conducted a comprehensive validation study on several diagnoses and procedures in the database. The agreement between hospital data and clinical charts was excellent for hysterectomy,^{50,61} ischemic heart disease,⁵¹ chronic obstructive pulmonary disease,⁵¹ and cholecystectomy.⁵⁰ When the researcher compared psychiatric diagnoses from several data sources, he found good agreement for schizophrenia but less so for depressive disorders.⁶⁴

Physician records have been used on a limited basis for validation. One study attempted to assess the usefulness of the databases for research on depression and its treatment.⁷⁹ For a cohort of antidepressant users, the three-digit ICD-9 diagnoses on the physician services database were compared to physicians' records. The agreement, sensitivity and specificity were good (77, 71, and 85% respectively).

WEAKNESSES

The health databases have been constructed by the Saskatchewan government primarily for program management purposes. Research, therefore, is a secondary use, and the databases may not be well suited to some types of study. As well, being administrative in nature, changes in policy and/or program features may influence the data collected. The gap in the drug database from July 1987 to December 1988 is an example of how a program

change affected the completeness of the data collected.

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 over 20 years of drug exposure data and almost 30 years of outcome/diagnostic data are available. Nevertheless, the databases still may be too small to evaluate low prevalence exposures or low incidence outcomes.

There are some limitations regarding exposure data. The Drug Plan operates on a formulary system. While the Saskatchewan Formulary is extensive (over 2800 drug products in 1998), the drug must be listed in the formulary or covered under a special program in order for records of its use to be included in the drug database. Consequently, it is not always possible to study exposure to prescription drugs as soon as they are available on the Canadian market. As well, there is no centralized computer database on inpatient drug use, over-the-counter drug use, or use of alternative therapies.

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. As well, 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 any epidemiological analyses, information on confounding is important. 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 record abstraction or patient interviews).

Some of these limitations may be offset by the ability to access individual patient records to obtain information not included in the computer databases. This is a manual process, however, and is therefore more time consuming and expensive.

PARTICULAR APPLICATIONS

Researchers interested in using the data may submit proposals to Saskatchewan Health. Policies and procedures are in place that enable the data to be used for research while maintaining beneficiary and provider confidentiality and database integrity. Researchers using the data have included those from academia, other governments, the pharmaceutical industry, other private companies, and practitioners.

The prime consideration for use of the databases is confidentiality of both recipients and providers of services. Identifiable information including, for example, the identity of the patient, physician, pharmacy, or hospital will not be released. As well, Saskatchewan Health will not allow source files, even if pseudo-identified, to be released; only aggregated information or limited fields of information at a person level may be released. Any data release must be consistent with the terms and conditions of Saskatchewan's Freedom of Information and Protection of Privacy Act. Further, information collected directly from study subjects will not be linked with the administrative data without the explicit consent of the study subjects.

The Research Services Unit of the Population Health Branch has responsibility for conducting and/or facilitating pharmacoepidemiology and other outcomes research, and is the point of contact for investigators who are interested in using the data for research. Personnel in the Unit have combined expertise in project management, research methodology, and analyses as well as extensive experience in using the large computerized databases for pharmacoepidemiology and epidemiology research.

The Saskatchewan databases have been used for a variety of studies.⁹⁻⁸¹ To illustrate the types of pharmacoepidemiology study that have been conducted, synopses of several projects are outlined below.

EXAMPLES OF USES

Studies of New Chemical Entities

Stang *et al.* conducted a study to determine the incidence of lactic acidosis in metformin users.⁷¹

This project was done in collaboration with the United States Food and Drug Administration (FDA) under a Cooperative Agreement.

Metformin, a biguanide oral antihyperglycemic agent, was a recent new chemical entity in the United States, but had been on the market in Canada since 1972 and listed in the Saskatchewan Formulary since 1981. The study design was historical cohort. Subject selection, based on information in the outpatient prescription drug database, identified 11 797 beneficiaries who received one or more prescriptions for metformin, for 22 296 person years of exposure. Within this cohort, ten cases were identified as having hospitalization with a primary, secondary, or other discharge diagnosis of acidosis (ICD-9: 276.2) within 120 days following the dispensing of a metformin prescription. Since the code 276.2 is not specific for lactic acidosis, but also includes metabolic, respiratory, and acidosis not otherwise specified, primary records for all cases were abstracted. An incidence rate of nine cases per 100 000 person years (95% confidence interval (0, 21)) was calculated based on the person years of exposure and two cases of lactic acidosis with substantiating lactate levels. One case had a possible spurious laboratory value which, if discounted, resulted in an incidence of 4.5 cases per 100 000 person years (95% confidence interval (0, 13)).

This was the first known population based longitudinal (16 years) postmarketing study with complete ascertainment of patients dispensed metformin and hospitalizations associated with lactic acidosis. Because the rate of lactic acidosis in metformin users is low, this study had the advantage of having enough accumulated person time of metformin exposure for detection of cases.

Studies of Drug Exposure and Health Outcomes

1. A study of the use of statin cholesterol lowering drugs and the risk of cancer is being conducted by Beck *et al.*^{73,76} The objectives of this study are to determine whether the use of any or all of the statins is associated with an increase in the overall risk of developing cancer or the risk of

developing certain types of cancer, in particular breast or gastrointestinal cancer.

A feasibility study was conducted to determine whether there were enough users of statins exposed to these drugs for a sufficiently long period to be able to detect an association between their use and the incidence of cancer. Based on the findings of the feasibility study, work on a historical cohort study is proceeding. Four study groups, defined based on their exposure or lack of exposure to certain lipid lowering drugs, are being identified for follow-up. The exposed cohort consists of those individuals identified as having received one or more prescriptions for statins during the period January 1989 to June 1997. For the breast cancer analyses, 13 592 female statin users were identified, with a total of 20 953 person years of exposure. The nonexposed comparison group was selected from the population of eligible beneficiaries who did not have any prescriptions for lipid lowering drugs during the study period. The remaining two comparison groups were selected based on their exposure to other commonly used lipid lowering drugs (cholestyramine or gemfibrozil). Use of these comparison groups will help to control for the presence of potentially confounding variables that may be associated with the use of lipid-lowering drugs (e.g., obesity).

2. A research project was undertaken by Pérez Gutthann *et al.* to look at the association between concurrent use of nonsteroidal anti-inflammatory drugs (NSAIDs) and several serious outcomes in the Saskatchewan population during the period 1982 to 1986. One of the outcomes of interest was acute renal failure.⁶³

The study population consisted of 228 392 patients who had received a prescription for a specific NSAID at any time from January 1982 to December 1986. Each study member was followed until admission to hospital for a renal disorder, death, emigration from Saskatchewan, or the end of the study period. Potential cases were initially identified from hospital separation data. As well, information was abstracted from hospital records by trained abstractors to validate cases and to obtain other

clinical information not available in the computerized databases. Controls were selected from the 228 392 members of the study population according to certain criteria. Person time of NSAID exposure was calculated and aggregated across calendar periods, sex, and five-year age strata for each of the NSAIDs into four main exposure categories (current users, recent users, past users, and nonusers). The researchers concluded that idiopathic acute renal failure is a rare occurrence among persons not using NSAIDs (about two cases per 100 000 person years), but that there was a fourfold increase in risk associated with NSAID use which was dose dependent and occurred especially during the first month of therapy.

The data compiled for this research project have also been used to study several outcomes related to gastrointestinal and liver effects.^{32-34,62,69}

Chronic Disease Surveillance and Burden Of Illness Studies

The Saskatchewan databases are being used to study diabetes in the province.^{72,75,77} All Saskatchewan Health beneficiaries who had one or more outpatient prescriptions for insulin or an oral hypoglycemic agent, two or more physician visits with a diagnosis of diabetes within a two year period, and/or one or more hospitalization for diabetes during the period January 1991 to December 1996 were identified. Annual incidence and prevalence rates were calculated.

A total of 45 716 individuals were identified as being diabetic in Saskatchewan during the study period. Of these, 52.1% were male and 86.8% were 40 years of age or older. The overall prevalence in 1996 was 3.7% (38 124 diabetics). Analyses are ongoing and will include descriptions of comorbidities and investigation of drug outcomes among the diabetic cohort. The dataset will be updated regularly to enable ongoing measurement of trends in diabetes in Saskatchewan. The general methodology will be applied to other chronic diseases for broader surveillance activities.

Adverse Drug Reaction Surveillance

A demonstration project using Saskatchewan's health databases was conducted to assess the feasibility of using administrative healthcare databases for adverse drug reaction (ADR) alerting.⁴¹

The drug database was used to identify the first 20 000 patients prescribed a new drug for the first time. These patients were then followed for a year after their initial prescription to identify all medical "events" experienced while on the medication. In this study, an "event" was defined as any diagnosis or medical service recorded in the physician services or inpatient hospital databases as well as any prescription recorded in the outpatient drug data. Two of the drugs selected for this project were nonsteroidal anti-inflammatory drugs: piroxicam and sulindac. The drugs were selected because their adverse effect profiles were reasonably well known and each drug had been covered by the Saskatchewan Drug Plan for an appropriate time period. Analyses examined age, sex, patterns of treatment continuation, and diagnostic "event" rates.

The investigator concluded that an acute ADR alerting system based on administrative data files like those in Saskatchewan is feasible and practical for the systematic evaluation of significant acute adverse effects of drugs. The cost of using such a system is relatively inexpensive and there are no followup problems.

Studies of Drug Use In Pregnancy

A study was carried out as a rapid response to a report from France of a highly significant association between maternal use of valproate for epilepsy during pregnancy and spina bifida in the offspring.¹² The purpose was to determine whether epileptic women were being prescribed valproate during pregnancy and, if so, whether use of the drug was associated with the occurrence of spina bifida or other malformations.

Initially the outpatient drug file was scanned for valproate prescriptions in 1980–81 for women aged 16–40. This identified 95 women. The drug records of the 95 valproate users were then linked to their hospitalization files via their unique

identification number for the period January 1980 to June 1982, and women who had a pregnancy related admission were identified. During the study period of interest, the Vital Statistics file did not use the unique identification number, so the file was searched by name for that period to validate the births found in the hospital file and also to identify stillbirths among the 95 women. Both sources found eight births among the valproate users, and no stillbirths. Finally, the outpatient drug histories were reviewed for the eight women to look more closely at the timing of the valproate prescriptions relative to the pregnancy.

In this study, the exposed population was too small to provide sufficient statistical power to detect an association. The project did, however, demonstrate the potential of using the Saskatchewan databases to study certain effects of drugs used during pregnancy. In addition, complete review of the records was finished within eight days of the initial decision to conduct the study, illustrating the value of relatively current and linkable health databases.

Economic Studies

1. Albright *et al.* used the Saskatchewan databases to assess the change in health care resource utilization and costs related to the initiation of risperidone therapy in patients with chronic schizophrenia.⁶⁵

The project was a retrospective cohort study that linked registry, prescription drug, hospital separation, physician services, and mental health services data. Study participants included all those patients who received at least one prescription for risperidone between July 1993 and December 1993. Utilization information from the databases was collected for equivalent periods (an average of 10 months) before and after initiation of risperidone therapy. The results demonstrated that treatment with risperidone produced substantial reductions in the utilization of health care services which translated to cost savings across hospital, physician, and mental health services.

The incremental cost of risperidone therapy was offset almost eightfold by these savings.

2. The Saskatchewan databases were used to examine information on care received and cost of that care as one component of a study of the cost-effectiveness of home care as a substitute for hospital care.⁸⁵

During the 1990s, hospital stays have become shorter and home care after hospital stays more common in Saskatchewan. The study was designed to provide health care providers and decision makers with information on whether this shift has affected patient health and whether it truly saves money. The results showed that about half of all adult medical and surgical patients could be discharged from hospital an average of two days sooner and that these patients could be provided with alternate care, mostly home care, to achieve the same health outcomes at less cost. The study also found that, while outcomes were the same, it costs more overall (\$830 CAN 1998) to provide a patient with nonacute care in hospital than to discharge them home with alternate followup care.

Drug and Other Health Service Utilization Studies

1. The use of cholesterol testing and drug treatment for lowering blood cholesterol have grown rapidly in the past decade. As part of a project to develop guidelines on cholesterol testing and treatment, the Saskatchewan Health Services Utilization and Research Commission (HSURC) examined adherence to cholesterol lowering agents in Saskatchewan using drug data from 1991 to 1993.⁵⁵ Analyses focused on 11 002 subjects defined as new users of lipid lowering drugs during the study period.

The researchers found that the proportion of new users stopping medication after filling only one prescription ranged from 30 to 38%. The estimated median time on medication for these new users was 111 days, and the proportion of new users remaining on medication at one year was 25%. Drug type was the single strongest predictor of time on drug therapy with the

pattern of adherence paralleling known drug side effects. For example, new users of niacin and cholestyramine were most likely to stop their medication, while those dispensed a statin were most likely to continue. The researchers concluded adherence to lipid lowering drug therapy in the Saskatchewan population was poor. After this study, the HSURC also examined more closely the factors that contribute to nonadherence.⁸⁶

2. Habbick *et al.* conducted a descriptive pharmacoepidemiology study. The prescription drug database was used to examine recent trends in use of inhaled β_2 -adrenergic agonists and inhaled corticosteroids and to determine whether these trends were in keeping with widely publicized guidelines recommending a reduction in the use of agents to treat symptoms (i.e., inhaled β_2 -adrenergic agonists) and increased use of prophylactic agents (i.e., inhaled corticosteroids).⁵²

All Saskatchewan residents 5 to 54 years of age who had received one or more prescriptions for a drug used to treat asthma (inhaled drugs and oral β_2 -adrenergic agonists and methylxanthines) from 1989 to 1993 were included. There was a steady increase in the proportion of the population receiving prescriptions for drugs to treat asthma over the five-year period. There was also a marked change in the patterns of prescriptions for drugs used to treat asthma with a shift toward the use of inhaled corticosteroids. In contrast, the proportion of patients who received β_2 -adrenergic agonists remained fairly constant. The authors concluded that the trends suggest many physicians adopted current guidelines advocating increased attention to prevention of airway inflammation in the treatment of asthma.

Product Assurance Evaluations

The databases have been used to provide information for "product assurance" agreements with the Saskatchewan Drug Plan.⁸⁷

When a new drug is introduced to the Canadian market, the manufacturer may submit a request to the Saskatchewan Drug Plan for the drug to be

considered for coverage. Part of the documentation submitted includes information on the manufacturer's expectations regarding utilization and health outcomes if the particular drug is listed in the formulary. If the Drug Plan's expert advisory committees recommend formulary listing for the drug, the Drug Plan may enter into a product assurance agreement with the manufacturer to evaluate whether the actual use of the particular drug after it is listed in the formulary is consistent with the prelisting information provided by the manufacturer. Depending on the differences between actual and expected use, an intervention may occur and a financial adjustment may be made against the manufacturer. The data (either drug data alone or in combination with other health services data) have been used to provide the information on the actual utilization and outcomes in Saskatchewan for such agreements.

THE FUTURE

One of the principles guiding future initiatives in the province's health system is the coordination and integration of health services for a more responsive, efficient, client-centred system. One of the objectives integral to this principle is the effective use and development of information, technology, planning, and reporting management systems.⁸⁸

Saskatchewan Health is developing new reporting systems to collect and provide data electronically to improve the accuracy, comprehensiveness, and timeliness of health information. For example, one information system under development will support the provincial immunization program. This system is expected to be functioning province-wide by April 1999. When fully implemented, it will operate through the Internet, and data will be collected from all public health offices in the province with infants and pre-schoolers being the priority. (A small proportion of vaccinations that are done in physicians' offices will not be included.) This database will be linkable with the other health databases and will enhance surveillance and other risk/benefit assessments.

As well, the Saskatchewan Health Information Network (SHIN) is being designed to build on Saskatchewan's existing fiber optic network and advances in information technology to establish computer connections between health service sites across the province. Through the information network, health service providers would be able to access the data when authorized, examine records and tests, and consult with clients and colleagues. The SHIN is still in early design stages, but the information technology developed will be valuable as it develops the information access using this technology: it will be valuable not only for direct patient care but also for population-based studies including research, using de-identified data, into the effectiveness of treatments and interventions.

Collection of appropriate information and preservation of confidentiality of personal health information will continue to be of paramount importance. In 1997 and 1998, Saskatchewan Health held extensive consultations within the province on the protection of personal health information. As a result of those consultations and in order to put a legal framework around good information practices, The Health Information Protection Act was passed in the legislature in the spring of 1999 with proclamation anticipated in 2000. This act will establish a common set of rules for everyone in the health system. It will emphasize protection of privacy of personal health information while ensuring that, with proper precautions, information is available to provide efficient health services. Research under specific conditions will continue to be a permitted use of health service information.

The value of the Saskatchewan health databases as a resource for pharmacoepidemiology research has been demonstrated. As enhancements of the databases continue and as the information is used in a responsible manner, the databases will continue to contribute to providing valuable information on public health issues related to drug use.

ACKNOWLEDGEMENTS

We gratefully acknowledge Dr Linda Strand both for her contribution to pharmacoepidemiology in

Saskatchewan and for co-authoring our chapter in the second edition of *Pharmacoepidemiology*.

We also thank our colleagues in Saskatchewan Health, in particular Margaret Baker, Jim Engel, Patrick Melia, and Valerie Phillips, for their review of the manuscript.

REFERENCES

1. Saskatchewan Health: International Symposium on Drug DataBase Uses, Proceedings. Regina, 1984.
2. Strand LM. Drug epidemiology resources and studies: the Saskatchewan database. *Drug Inf J* 1985; **19**: 253–6.
3. Tilson H. Getting down to bases: record linkage in Saskatchewan [editorial]. *Can J Public Health* 1985; **76**: 222–3.
4. West R. Saskatchewan Health: A developing resource. *Am J Prev Med* 1988; **4** (suppl 2): 25–7.
5. Strand LM, West R. Health data bases in Saskatchewan. In: Strom B, ed., *Pharmacoepidemiology*. New York: Churchill Livingstone, 1989; 189–200.
6. Rawson NSB, D'Arcy C, Blackburn JL, Bowen RC, Wallace SM, Downey W, Senft D. *Epidemiologic Research Using Linked Computerized Health Care Datafiles in Saskatchewan, Canada*, Technical Report 2. Saskatoon, Saskatchewan: Psychiatric Pharmacoeepidemiology Research Consortium; 1992.
7. Malcolm E, Downey W, Strand LM, McNutt M, West R. Saskatchewan Health's linkable databases and pharmacoepidemiology. *Post-Marketing Surveillance* 1993; **6**: 175–264.
8. Strand LM, Downey W. Health Databases in Saskatchewan. In: Strom B, ed., *Pharmacoepidemiology*, 2nd edn. Chichester: Wiley, 1994; 217–29.
9. Skoll EL, August RJ, Johnson GE. Drug prescribing for the elderly in Saskatchewan during 1976. *Can Med Assoc J* 1979; **121**: 1074–81.
10. Schnell BR. A review of the use of prescription drugs in Saskatchewan. *Can Pharm J* 1981; **7**: 267–70.
11. Power B, Downey W, Schnell BR. Utilization of psychotropic drugs in Saskatchewan. *Can J Psychiatry* 1983; **28**: 547–51.
12. West R, Sherman GJ, Downey W. A record linkage study of valproate and malformations in Saskatchewan. *Can J Public Health* 1985; **76**: 226–8.
13. Quinn DMP. Prevalence of psychoactive medication in children and adolescents. *Can J Psychiatry* 1986; **31**: 575–80.
14. Guess HA, West R, Strand LM, *et al*. Hospitalization for renal impairment among users and non-users of nonsteroidal anti-inflammatory drugs in Saskatchewan, Canada, 1983. In: Rainsford KD, Velo GP, eds, *Side-Effects of Anti-Inflammatory Drugs, Part 2: Studies in Major Organ Systems*. Lancaster: MTP Press, 1987; 367–74.
15. Babiker IE. Comparative efficacy of long-acting depot and oral neuroleptic medications in preventing schizophrenic recidivism. *J Clin Psychiatry* 1987; **48**: 94–7.
16. Hogan DJ, Strand LM, Lane PR. Isotretinoin therapy for acne: a population-based study. *Can Med Assoc J* 1988; **138**: 47–50.
17. Guess HA, West R, Strand LM, Helston D, Lydick EG, Bergman U, Wolski K. Fatal upper gastrointestinal hemorrhage or perforation among users and nonusers of nonsteroidal anti-inflammatory drugs in Saskatchewan, Canada 1983. *J Clin Epidemiol* 1988; **41**: 35–45.
18. Ray WA, Griffin MR, Downey W, Melton LJ. Long-term use of thiazide diuretics and risk of hip fracture. *Lancet* 1989; **1**: 687–90.
19. Ray WA, Griffin MR, Downey W. Benzodiazepines of long and short elimination half-life and the risk of hip fracture. *J Am Med Assoc* 1989; **262**: 3303–7.
20. Rothman KJ, Funch DP, Dreyer NA. Bromocriptine and puerperal seizures. *Epidemiology* 1990; **1**: 232–8.
21. Ray WA, Griffin MR, Malcolm E. Cyclic anti-depressants and the risk of hip fracture. *Arch Intern Med* 1991; **151**: 754–6.
22. Gormley M, Barer M, Melia P, Helston D. *The Growth in Use of Health Services*. Regina, Saskatchewan: Saskatchewan Health, 1989.
23. Blackburn JL, Downey W, Quinn TJ. The Saskatchewan program for rational drug therapy: effects on utilization of mood-modifying drugs. *Drug Intell Clin Pharm* 1990; **24**: 878–82.
24. Thiessen BQ, Wallace SM, Blackburn JL, *et al*. Increased prescribing of antidepressants subsequent to beta-blocker therapy. *Arch Intern Med* 1990; **150**: 2286–90.
25. O'Reilly R. Are the monoamine oxidase inhibitors facing extinction? *Can J Psychiatry* 1991; **36**: 186–9.
26. Horwitz RI, Spitzer W, Buist S, Cockcroft D, Ernst P, Habbick B, *et al*. Clinical complexity and epidemiologic uncertainty in case-control research: fenoterol and asthma management. *Chest* 1991; **100**: 1586–91.
27. Suissa S, Spitzer WO, Abenhaim L, Downey W, Gardiner RJ, Fitzgerald D. Risk of death from human insulin. *Pharmacoepidemiol Drug Safety* 1992; **1**: 169–75.
28. Quinn K, Baker M, Evans B. Who uses prescription drugs? Results from a population-wide study in Saskatchewan. Regina, Saskatchewan: Saskatchewan Health; 1992.
29. Quinn K, Baker M, Evans B. A population-wide profile of prescription drug use in Saskatchewan, 1989. *Can Med Assoc J* 1992; **146**: 2177–86.

30. 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. *New Engl J Med* 1992; **326**: 501–6.
31. Ernst P, Spitzer WO, Suissa S, Cockcroft D, Habbick B, Horwitz RI, *et al.* Risk of fatal and near-fatal asthma in relation to inhaled corticosteroid use. *J Am Med Assoc* 1992; **268**: 3462–3464.
32. García Rodríguez LA, Walker AM, Pérez Gutthann S. Nonsteroidal anti-inflammatory drugs and gastrointestinal hospitalizations in Saskatchewan: a cohort study. *Epidemiology* 1992; **3**: 337–42.
33. García Rodríguez LA, Pérez Gutthann S, Walker AM, Lueck L. The role of nonsteroidal anti-inflammatory drugs in acute liver injury. *Br Med J* 1992; **305**: 865–8.
34. Pérez Gutthann S, García Rodríguez LA. Increased risk of hospitalizations for acute liver injury in a population with exposure to multiple drugs. *Epidemiology* 1993; **4**: 497–501.
35. Malcolm E, Downey W, Strand L, Rawson NSB. Ongoing measurements of data quality within Saskatchewan's linkable health databases [abstract]. *Post-Marketing Surveillance* 1992; **6**: 42–3.
36. 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.
37. Tennis P, Andrews E, Bombardier C, Wang Y, Strand L, West R, *et al.* Record linkage to conduct an epidemiologic study on the association of rheumatoid arthritis and lymphoma in the province of Saskatchewan, Canada. *J Clin Epidemiol* 1993; **46**: 685–95.
38. Ernst P, Habbick B, Suissa S, Hemmelgarn B, Cockcroft D, Buist AS, *et al.* Is the association between inhaled beta-agonist use and life-threatening asthma because of confounding by severity? *Am Rev Respir Dis* 1993; **148**: 75–9.
39. Habbick BF, for the Saskatchewan Asthma Epidemiology Project Team. The Saskatchewan Asthma Epidemiology Project. *Sask Med J* 1993; **4**: 4–7.
40. Suissa S, Hemmelgarn B, Spitzer WO, Brophy J, Collet JP, Cote R, *et al.* The Saskatchewan oral contraceptive cohort study of oral contraceptive use and cardiovascular risks. *Pharmacoepidemiol and Drug Safety* 1993; **2**: 33–49.
41. Rawson Nigel SB. An acute adverse drug reaction alerting scheme using the Saskatchewan Health datafiles. *Drug Invest* 1993; **6**: 245–56.
42. Risch HA, Howe GR. Menopausal hormone usage and breast cancer in Saskatchewan: a record-linkage cohort study. *Am J Epidemiol* 1994; **139** (7): 670–83.
43. Downey W, Malcolm E, Strand LM, McNutt M. Pharmacoepidemiology: an example of exposure/outcome health research. In: McDuffie H, *et al.*, eds, *Supplement to Agricultural Health and Safety: Workplace, Environment, and Sustainability*. Saskatoon: Centre for Agricultural Medicine, 1994; 157–164.
44. Lichtenstein MJ, Griffin MR, Cornell JE, Malcolm E, Ray WA. Risk factors for hip fractures occurring in the hospital. *Am J Epidemiol* 1994; **140**: 830–8.
45. Suissa S, Ernst P, Boivin JF, *et al.* A cohort analysis of excess mortality in asthma and the use of inhaled beta-agonists. *Am J Respir Crit Care Med* 1994; **249**: 604–10.
46. Hemmelgarn B, Blais L, Collet, JP, Ernst P, Suissa S. Automated databases and the need for fieldwork in pharmacoepidemiology. *Pharmacoepidemiol Drug Safety* 1994; **3**: 275–82.
47. Suissa S, Blais L, Ernst P. Patterns of increasing β -agonist use and the risk of fatal or near-fatal asthma. *Eur Respir J* 1994; **7**: 1602–9.
48. Ernst P, Hemmelgarn B, Cockcroft DW, Suissa S. Overreliance on bronchodilators as a risk factor for life-threatening asthma. *Can Respir J* 1995; **2**: 34–9.
49. Tennis P, Cole TB, Annegers JF, Leestma JE, McNutt M, Rajput A. Cohort study of incidence of sudden unexplained death in persons with seizure disorder treated with antiepileptic drugs in Saskatchewan, Canada. *Epilepsia* 1995; **36**: 29–36.
50. Rawson NSB, Malcolm E. *Validity of the Recording of Cholecystectomy and Hysterectomy in the Saskatchewan Health Care Datafiles*, Technical Report 3. Saskatoon, Saskatchewan: Pharmacoepidemiology Research Unit; 1995.
51. 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.
52. Habbick B, Baker MJ, McNutt M, Cockcroft DW. Recent trends in the use of inhaled β_2 -adrenergic agonists and inhaled corticosteroids in Saskatchewan. *Can Med Assoc J* 1995; **153**: 1437–43.
53. Neutel I, Downey W, Senft D. Medical events after a prescription for a benzodiazepine. *Pharmacoepidemiol and Drug Safety* 1995; **4**: 63–73.
54. Neutel CI. Risk of traffic accident injury after a prescription for a benzodiazepine. *Ann Epidemiol* 1995; **5**: 239–44.
55. Health Services Utilization and Research Commission. *Cholesterol Testing and Treatment in Adults. Final Report*. Saskatoon: HSURC, 1995.
56. Joseph KS, Blais L, Ernst P, Suissa S. Increased morbidity and mortality related to asthma among asthmatic patients who use major tranquilizers. *Br Med J* 1996; **312**: 79–83.
57. Suissa S. The case–time–control design. *Epidemiology* 1996; **6**: 248–53.
58. Suissa S, Hemmelgard, Blais L, Ernst P. Bronchodilators and acute cardiac death. *Am J Respir Crit Care Med* 1996; **154**: 1598–602.

59. 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.
60. Raiford DS, Pérez Gutthann S, García Rodríguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. *Epidemiology* 1996; **7**: 101–4.
61. Edouard L, Rawson NSB. Reliability of the recording of hysterectomy in the Saskatchewan health care system. *Br J Obstet Gynaecol* 1996; **103**: 891–7.
62. Pérez Gutthann S, García Rodríguez LA, Raiford DS. Individual nonsteroidal anti-inflammatory drugs and other risk factors for upper gastrointestinal bleeding and perforation. *Epidemiology* 1996; **8**: 18–24.
63. Pérez Gutthann S, García Rodríguez LA, Raiford DS, Duque Oliart A, Ris Romeu J. Nonsteroidal anti-inflammatory drugs and the risk of hospitalization for acute renal failure. *Arch Internal Med* 1996; **156**: 2433–9.
64. Rawson NSB, Malcolm E, D'Arcy C. Reliability of the recording of schizophrenia and depressive disorder in the Saskatchewan health care datafiles. *Soc Psychiatry Psychiatr Epidemiol* 1997; **32**: 191–9.
65. 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.
66. Neutel CI, Maxwell CJ. The benzodiazepine treadmill: does one prescription lead to more? *Pharmacoepidemiol Drug Safety* 1996; **5**: 39–42.
67. Remillard AJ. A pharmacoepidemiological evaluation of anticholinergic prescribing patterns in the elderly. *Pharmacoepidemiol Drug Safety* 1996; **5**: 155–64.
68. Johnson JA, Wallace SM. Investigating the relationship between β -blocker and antidepressant use through linkage of the administrative databases of Saskatchewan Health. *Pharmacoepidemiol Drug Safety* 1997; **6**: 1–11.
69. Pérez Gutthann S, García Rodríguez LA, Raiford DS. Individual non-steroidal anti-inflammatory drugs and hospitalizations for upper gastrointestinal bleeding and perforation in Saskatchewan: A nested case-control study. *Epidemiology* 1997; **8**: 18–24.
70. Maxwell CJ, Neutel CI, Hirdes JP. A prospective study of falls after benzodiazepine use: a comparison of new and repeat use. *Pharmacoepidemiol Drug Safety* 1997; **6**: 27–35.
71. Stang MR, Lueck L, Wysowski D, Downey W, Butler-Jones D. Meformin and lactic acidosis [abstract]. *Can J Clin Pharmacol* 1997; **5**: 53.
72. Beck P, Downey W, Osei W, Baker M. Use of administrative data to create disease-specific datasets for disease surveillance and program planning [abstract]. *Can J Clin Pharmacol* 1998; **5**: 54.
73. Beck P, Downey W, Butler-Jones D, et al. Use of statins and the risk of breast cancer [Abstract]. *Can J Clin Pharmacol* 1998; **5**: 54.
74. Beck P, Downey W, Butler-Jones, Lueck L. Pharmaceutical costs of patient non-adherence to lipid-lowering drug therapy [abstract]. *Value Health* 1998; **1**: 50.
75. Beck P, Downey W, Osei W, Baker M. Use of administrative data to create disease-specific datasets for use in health economic studies [abstract]. *Value Health* 1998; **1**: 72.
76. Beck P, Downey W, Butler-Jones D, et al. Use of statins and the risk of breast cancer. *Pharmacoepidemiol Drug Safety* 1998; **7** (suppl 2): S88.
77. Beck P, Downey W, Osei W, Baker M. Use of administrative data to create disease-specific datasets for disease surveillance and program planning. *Pharmacoepidemiol Drug Safety* 1998; **7** (suppl 2): S86.
78. Senthilselvan A. Prevalence on physician-diagnosed asthma in Saskatchewan 1981 to 1990. *Chest* 1998; **114**: 388–92.
79. West SL, Richter A, Nennstiel ME, Mauskopf JA. *Saskatchewan Database Assessment: Draft Report*. Indianapolis, IN: Research Triangle Institute, for Eli Lilly and Company, 1998.
80. Caro JJ, Salas M, Speckman JL, Raggio GD, Jackson JD. Persistence with treatment for hypertension in actual practice. *Can Med Assoc J* 1999; in press.
81. Caro JJ, Salas M, Speckman JL, Raggio GD, Jackson JD. The importance of actual practice data for rational clinical decisions: initial choice of medication in hypertension. *Can Med Assoc J* 1999; in press.
82. Health Registration. *Covered Population 1998*. Regina: Saskatchewan Health, 1998.
83. World Health Organization. *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death, 9th Rev*. Geneva: World Health Organization, 1977.
84. Statistics Canada. *Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures*. Ottawa, Ontario: Minister of Supply and Services, 1986.
85. Health Services Utilization and Research Commission. *Hospital and Home Care Study*. Summary Report 10. Saskatoon: HSURC, 1998.
86. Health Services Utilization and Research Commission. *Adherence to Cholesterol-Lowering Drugs in*

- Saskatchewan. Summary Report 9.* Saskatoon: HSURC, 1997.
87. Drug Plan and Extended Benefits Branch, Saskatchewan Health. *Annual Statistical Report 1996/97.* Regina: Government of Saskatchewan, 1997.
88. Saskatchewan Health. *Health System Directions. Part II: Objectives and Strategies.* Regina: Saskatchewan Health, 1998.

21

Automated Pharmacy Record Linkage in The Netherlands

HUBERT G. LEUFKENS

Department of Pharmacoepidemiology and Pharmacotherapy, Universiteit Utrecht, Utrecht, The Netherlands

JOHN URQUHART

Department of Epidemiology, Maastricht University, Maastricht, The Netherlands, and AARDEX Ltd., Zug, Switzerland, and Palo Alto, CA, USA

INTRODUCTION

Pharmacoepidemiology is a relatively new discipline in The Netherlands but, as in other countries, it has many antecedents, each with some decades of history. The traditions of academic medicine in The Netherlands are long and strong. Bedside teaching as a basic pedagogic device was pioneered by Herman Boerhaave in Leiden in the early 18th century, attracting students from as far as Japan. Around the turn of this century, van't Hoff's discovery of osmosis and Einthoven's discovery and development of electrocardiography laid important scientific foundations for modern medicine; both were early Nobel Laureates. Clinical research developed in The Netherlands in parallel with other Western countries, with increasing use of randomized clinical trials after the Second World War.

In the latter 1980s, a number of Dutch clinical pharmacologists began shifting towards observational studies, bridging the gap between clinical

pharmacology and epidemiology. Notable findings in this work have been, for example, the role of dietary potassium intake in prevention of cardiovascular disease,¹ information about the risks of stroke in patients on antihypertensive treatment,² and the adverse effects on the cardiovascular system of the use of nonsteroidal anti-inflammatory drugs (NSAIDs).³ The latter study has stimulated further pharmacoepidemiologic research with the PHARMO-system (more about this later) on the association between NSAID use and adverse cardiovascular effects. Heerdink *et al.* found a twofold increased risk of hospitalization for CHF during periods of concomitant use of diuretics and NSAIDs compared with use of diuretics only.⁴

As elsewhere, the long term consequences of oral contraceptive use became a focus of what we would today call pharmacoepidemiology research, with a series of studies investigating the preventive effect of oral contraceptives on the risk of rheumatoid arthritis,^{5,6} followed more recently

by studies from the same group on mechanisms and magnitude of thrombogenic risk with so-called third generation oral contraceptives.^{7,8} The early oral contraceptive studies, together with those performed at Mayo Clinic, have contributed greatly to the awareness of methodologic problems of case-control studies.⁹

Two notable antecedents of Dutch pharmacoepidemiology are pharmacovigilance and drug utilization research. The Dutch pharmacovigilance system began in the 1960s, and has generated a succession of signals of suspected adverse effects, generating hypotheses and fostering pharmacoepidemiology studies to test them¹⁰⁻¹⁴ (see also Chapter 11). Drug utilization research dates back to the early 1970s under the WHO Drug Utilization Research Group, recently renamed EURO-DURG, strengthening methods for ascertaining drug exposure^{15,16} (see also Chapter 29). These two lines of research have been strongly reinforced by the essentially universal computerization of Dutch community pharmacies in the mid-1980s, which has played a key role in a cornerstone of Dutch pharmacoepidemiologic studies: the reliable ascertainment of drug exposure.¹⁷

This focus on the reliable ascertainment of drug exposure anticipates the future integration of a strong pharmacodynamic perspective into the analysis of pharmacoepidemiology data, for the very basic reason that the actions of drugs depend on both dose and dose timing. Temporal sequence is basic to the inference of causality, as causes precede but do not follow effects. To exploit this principle in pharmacoepidemiology, requires reliable information on the time sequence of events in relation to initial and refill dispensing of drugs. When the record of events is scattered over multiple pharmacies, physicians, and hospitals, it becomes virtually impossible to give adequate descriptions of drug exposure, clinical correlates, and their temporal relations with drug exposure. Thus, in considering the importance of having reliable answers to the questions "which patients got which drugs, in what doses, and when?," one cannot overemphasize the "s" in "drugs," given that multiple drugs are the rule, not the exception, in much of pharmacotherapy.¹⁷

Progress in pharmacoepidemiology in The Netherlands has, as elsewhere, been closely related to specific features of the systems for healthcare delivery and its reimbursement. The discipline could flourish in environments where there is good access to data on drug exposure, diagnoses, reasons for hospitalization, clinical status, and outcomes. Yet there is no Valhalla where all these sources of information freely come together, so the possibilities for pharmacoepidemiology research vary greatly from one country to another, and within countries, as the different chapters in this section of this book demonstrate. Indeed, differences in the organization and reimbursement of healthcare that are trivial from an administrative perspective can have major impacts on the possibilities for epidemiologic research—a point that epidemiologists have traditionally found difficult to communicate to administrators. It was only a decade ago that Ellwood underscored the value of administrative and reimbursement arrangements that facilitate the collection of data on the vast array of natural experiments that comprise routine healthcare for outcomes research.¹⁸ Building a bridge between (clinical) epidemiology and outcomes research based on such data from routine health care remains a major challenge. There are still ample opportunities for happenstance, serendipity, ingenuity, and individual initiative to allow individual researchers to glean useful information from reimbursement systems that had been designed to fit local customs and administrative convenience.¹⁹

DESCRIPTION

DUTCH MEDICINE

Pharmacoepidemiology in The Netherlands is naturally imbedded in the principles of medicine as practiced in this country of approximately 15 million people. Demographically, The Netherlands is similar to the other Western countries—very low mortality rates until past age 60, with an increasing population of elderly, in which females outnumber males. The three main causes of death are, as in other technologically advanced

countries, coronary heart disease and other cardiovascular diseases, the various cancers, and accidents. *Per capita* expenditures on health care are in line with other western European countries, and about half of those in the United States. One searches in vain, however, for substantive differences in public health indices between The Netherlands and the United States.

The drug regulatory system in The Netherlands has been widely acknowledged as one of the most stringent in Europe, and plays an important role in European Regulatory Affairs within the framework the European Agency for the Evaluation of medicinal products (EMEA) in London. Governmental and reimbursement policies in The Netherlands show a certain reluctance towards new pharmaceutical products. Rising health care costs reinforce this policy. As a result, and because of the relatively small size of the country, the adoption of new drugs by Dutch prescribers leads only exceptionally to more than 8000 recipients of a new drug within its first year in the market. Thus, for most new drugs, The Netherlands is too small for the detection of very rare events.

Pharmaceutical prices in The Netherlands have long been among the highest in Europe, although recent regulatory action has forced industry to harmonize pricing according to European Union (EU) levels.

Dutch medicine is closely connected to British and North American medicine, contributing substantially to British and American medical journals, monographs, and textbooks. This fact is in keeping with the strongly international outlook that pervades many aspects of Dutch economic and social life. It is reflected in the national proficiency in English, German, and French languages, and is a natural consequence of the country's need to compete successfully in international markets. There are no special conceptions of disease, as one finds, for example, in the Teutonic conception of hypotonia or the hepatocentricity of Gallic medicine. If there is any special feature to Dutch medicine, it is a much more liberal use of oral anticoagulants in clinical situations where the risk of intravascular thrombosis is elevated. All patients prescribed oral anticoagulants in The Netherlands have their anticoagulant dosing man-

aged by a specially organized Thrombosis Service, which focuses exclusively on this difficult area of ambulatory pharmacotherapy. This specialization is believed to tip the risk-benefit balance much further in favor of benefit than in other countries where anticoagulation is in the hands of GPs, internists, or cardiologists. Moreover, this system has proved to provide an excellent source for tracing patients with bleeding disorders for epidemiologic research.²⁰

The Dutch system of medicine is based on primary care physician—general practitioners (GPs) called “house doctors” (*huisarts*)—who practice in the community but not in hospitals, referring ambulatory patients to specialists for out- or inpatient care, as circumstances require. Hospital care, as in other continental countries, is provided by full time staff physicians who are specialists of various kinds. Medical care, including prescription drugs, is essentially fully paid for by public or private insurers. Hospitals are organized regionally, with well defined catchments. It is, in principle, possible to link data from the community and hospitals, though with strict attention to anonymization, as discussed later. Health care of essentially uniform quality is provided to all citizens. Neither Dutch medicine nor the Dutch pharmaceutical industry is much influenced by the tort system of medical malpractice and product liability that is so prominent in the United States, and increasing in the United Kingdom.

The public insurers are regional agencies collectively called *Ziekenfonds* (“Sickfunds”). They provide essentially complete insurance coverage for approximately two-thirds of the population who fall below a defined annual income level. Patients covered by Sickfunds are required to designate a general practitioner for primary care and a community pharmacy for all reimbursed prescription drugs. The remaining third of the population are covered by private insurers, usually organized through one's employer. All insurers are nominally in competition with one another, but it is a muted form of competition that has, thus far, avoided serious poaching on each other's actuarial bases.

Like other social democracies in Western Europe, the Dutch have tax policies that tend to

minimize disparities in income. These socio-economic factors have important corollaries for health care. It would, for example, be unthinkable for a Dutch physician to decline to treat Sickfunds-insured patients. Inevitably there will be reflections of differential affluence in health data from Sickfunds-insured and privately insured, but minimization of the socio-economic and health factors that would underlie such differences has been one of the main driving forces in Dutch political and social life throughout this century.²¹ However, what the future will bring remains uncertain. The interplay between demographic and socio-economic factors, technological change, and other forces in the health care market, may result in very different, but all plausible scenarios.²²

GENERAL PRACTICE SYSTEM

Computerization of general practices began in the late 1980s. Debate and confusion about taxonomy and software have impeded its development, as has the lack of any distinct economic incentive for physicians to abandon paper based records. This situation contrasts sharply with the almost complete computerization of community pharmacies during the 1980s, facilitated by distinct economic incentives, a well developed taxonomy for products, and major investment in the development and marketing of several competing systems of software.

As in the United Kingdom, various schemes of pharmaceutical promotion, thinly veiled as post-marketing surveillance studies, succeeded in placing microcomputers in the offices of many Dutch physicians. These schemes eventually attracted attention from news media and professional organizations, with controversy and a certain overreaction that has cast suspicion on all studies done with marketed products. Whatever the extent of "seeding" of new pharmaceuticals these schemes may have achieved, they certainly succeeded in "seeding" computers among physicians, but it is still not clear that these schemes have brought computers and computer skills to a critical mass of Dutch physicians. Meanwhile, each year's new crop

of medical graduates bring their computer skills into the profession. Another aspect of "seeding" types of promotional activity is discussed later.

Since the late 1970s, a number of research networks of "sentinels" in general practice have been established among primary care physicians, doing population based (pharmaco)epidemiology studies. For example, Hoes conducted a case-control study among hypertensive patients, demonstrating an association between the use of potassium wasting diuretics and an increased risk of sudden cardiac death, based on data collected through a network of GPs in Rotterdam.²³ Pharmacy data were used to validate the drug exposure data provided by the GPs. No significant discrepancies between the two sets of data were found, which is noteworthy, in view of numerous examples of incomplete prescribing information from paper-based GP records.

Visser *et al.* looked at the association between the use of ACE inhibitors and cough in a network of the practices of 161 Dutch GPs, in which all consultations, morbidity, mortality, medical interventions, and prescriptions were registered (the NIVEL study).²⁴ A case-control study in this population showed that female cases of cough were three times more likely to be exposed to ACE inhibitors than noncough controls. This association was not seen in men.

An interesting development is seen in 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 system has been developed by the Department of Medical Informatics of the Erasmus University Medical School. The database included all demographic information, patient complaints, symptoms and diagnosis (ICPC), laboratory tests, discharge and consultant letters, and detailed prescription information (drug name, ATC code, dosing information, and indication). The system covers over 400 000 people, is population based, and provides a powerful tagging system enabling prospective followup of well defined target groups.^{25,26}

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 8000–14 000 patients per pharmacy. Dutch pharmacies essentially limit themselves to prescription drugs, and have none of the long shelves of consumer products that one finds in pharmacies in the United States, Canada, and the United Kingdom. Dutch pharmacies do carry over-the-counter products, but keep them behind the counter. In the past five years, the government has acted in various ways to restrict pharmacists' incomes, but most Dutch community pharmacies continue to be economically strong. Computerization of pharmacy records, and thus the compilation of prescription drug histories, is almost universal. Sickfunds insured patients, as noted above, receive all their reimbursed prescription drugs from a designated community pharmacy. The remaining third of the population, with private insurance, have the option to use more than one pharmacy—a frequent occurrence in cities, but infrequent and usually identifiable among those who reside in villages, which often have only one pharmacy. Thus, there is ample opportunity to identify complete ambulatory prescription drug histories in these pharmacies' computers, even among privately insured patients. In rural areas, an occasional dispensing physician is found, but their numbers are decreasing due to governmental policies.

Grouping patients as recipients of a particular pharmaceutical is thus possible by linking medication files, creating a pharmacy based cohort, as first defined by Borden at Upjohn in the 1970s.²⁷ Borden's method consisted of the identification and followup of patients whose treatment with a certain pharmaceutical is defined by pharmacy records showing that the product in question had been dispensed to them. After obtaining informed consent from the patient, he collected data on health outcomes by means of written questionnaires and telephone followup. Borden's approach has now been widely adopted. For example,

recipients of acitretine in The Netherlands were traced nationally through pharmacies and dispensing physicians, after it became known that the teratogen etretinate was a metabolite of acitretine. A written survey to all pharmacies and dispensing physicians was responded to by 87% of all dispensers, representing the same proportion of recipients of etretinate.²⁸ Using this unique source of drug exposure information, several networks of so-called "sentinel" pharmacies have been developed for pharmacoepidemiology research.²⁹ More on this will follow.

MEDICAL RECORD LINKAGE

There is widespread, deep seated resistance in The Netherlands to the use of a unique personal identification number. Identification for insurance purposes is done by an anonymized family number, with substring codes used to distinguish individuals within the family, to enable linkage of individual patients' 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. Coded and sealed envelopes were used for the exchange of information, maintaining the patients' anonymity but allowing correlation of drug exposure information with the corresponding patients' medical data. In one of these studies, however, the response rate of the GPs was only 81%, which could be a substantial obstacle to certain study designs.³⁰ Moreover, the procedure was time consuming and costly. The same sealed envelope scheme was used by Petri *et al.* to search for clinical signs of Parkinson's disease in recipients of flunarizine.³¹

In the early 1990s, a formal system of record linkage 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 patients' birth date, gender, and GP code, preserving anonymity.³² While there are certain probabilistic aspects of the linkage, the combination of

these three data items yields a sensitivity and specificity of over 95% each. The PHARMO system has been expanded to a population of over 500 000, is population based, links all prescription drug data to hospital data,^{4,33,34} and has been used to study a number of drug effects severe enough to require hospitalization. Presently, PHARMO is being linked also to primary care data, population surveys, cancer and accident registries, mortality data, and economic outcomes. In the absence of a unique patient identifier—which remains lacking in The Netherlands—the PHARMO system provides a powerful approach for conducting follow-up studies, case-control studies, and other analytical epidemiologic studies for evaluating drug induced effects. The data collection is longitudinal and goes 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. PHARMO figures in a number of applications discussed below.

STRENGTHS

The most important advantage of the Dutch system lies in its complete coverage of a relatively homogeneous population of reasonable size. We believe that the quality of data on drug exposure in The Netherlands is second to none, because of three factors: (i) computerized dispensing records are subject to financial audit because they are the basis for reimbursement; (ii) a long tradition that patients frequent a single GP and pharmacy; and (iii) the lack of economic incentive for between pharmacy shopping. Data on drug exposure can be linked to various outcomes and to data on possible confounders or effect modifiers. 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.^{7,35}

A second major advantage are the strong liaisons between patients, physicians, and pharmacists, facilitating the gathering of crucial followup information. One expression of these liaisons is a regional adverse drug reactions reporting scheme,

called LAREB (Landelijke Regionale Evaluatie Bijwerkingen), based on liaison between pharmacists and physicians.³⁶ LAREB plays a key role in pharmacovigilance, in close collaboration with the national regulatory agencies. The system provides prompt feedback to reporting physicians, aiming thereby to increase the interest in, and incentive for, ADR reporting. A rather more indirect approach is represented by a national scheme for collaboration between physicians and pharmacists, called Pharmacotherapy Discussion Groups.³⁷ These are monthly meetings between a pharmacist and his/her main prescribers—rarely more than ten physicians—to discuss a pertinent topic in pharmacotherapeutics. A national foundation provides programmatic teaching materials for these small group meetings, on the premise that the meetings improve the general level of understanding of prescription drugs, reinforce rational prescribing, and motivate the recognition and reporting of adverse drug reactions.

WEAKNESSES

The Dutch system is not immune from problems, such as unreliable or outdated information in the patient file undermining the quality of pharmacy records. Patient files in pharmacies are maintained mostly by hand, 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. The occurrence of missing data on gender or birth date, and probably fictitious birthdates, e.g., the frequently encountered 01–01–01 or 11–11–11, exemplify the concern. Today, a number of pharmacy computer systems include logical checking procedures in order to minimize the entry of false data in patient files.

There is a growing need for quality assessment of data for epidemiology research that are derived from medical information systems. As in many countries, the patient's medical record, being in the past a personal registry compiled by the individual health professional, has become a target for numerous interest groups, e.g., insurance compa-

nies, regulators, consumer organizations, lawyers involved in liability actions, and epidemiologists. 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. For reasons already discussed, pharmacy dispensing data are of generally high quality. However, proper evaluation of the quality and validity of data is always needed.³⁸⁻⁴⁰

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. Yet there is a strong sense that health care costs must be reduced, which has led to a series of recent changes that, from the perspective of the situations in health care in other countries, suggest failure to heed the warning "if it ain't broke, don't fix it." In a certain sense, we suffer the consequences of having a system of health care that evolved before extensive, effective interventions were possible. Effective interventions are the dividends of the large investment in biomedical research that all advanced nations have made since the end of the Second World War. So, interventions that "work" have grown in number, complexity, and cost, with steeply accelerating growth during the past two decades, but they have arrived piecemeal, changing the face of medicine in the fashion of a complex picture puzzle only gradually coming together, unguided by a view of the completed picture. The political and economic reactions have been similarly piecemeal, naturally lagging behind the technological changes, as the political and economic reactions develop in response to, never in anticipation of, accumulating problems created by growing costs, ethical dilemmas, and—the ultimate dividend—demographic changes.

There is also pressure to stimulate competition among pharmacies by freeing Sickfunds patients from the obligation to designate a single pharmacy. Yet Dutch patients have no economic incentive to use one pharmacy over another. Thus, personal service is the main basis for competition among pharmacies. It remains to be seen whether provision of "pharmaceutical care," as currently conceptualized, will develop into a competitive force.

PARTICULAR APPLICATIONS

THE DIMENSION OF TIME IN DRUG EXPOSURE

The ascertainment of drug exposure is determined by various factors. Time is a crucial variable that has the power to reject hypotheses about causality, e.g., when the putative adverse reaction precedes the first record of dispensing of the suspected drug. Thus, in assessing causal relationships between drug use and suspected events, it is essential to verify the temporal sequence of events in relation to exposure to the drug. This very basic point is not an easy one to apply, because of the prevalence of timing errors in casual reporting: an ostensibly minor error in data entry can transpose the onset of drug exposure, or the onset of disease, negating or supporting a hypothesis about causality.^{41,42}

A time oriented database consists essentially of three distinct types of variable: headers, point events, and interval events. Interest focuses not only on the interval between the beginning and end of drug exposure, termed the "risk window," but on periods prior to, and after, the risk window, for these provide information on events in the patient's life at times proximate to, but outside, the risk window. When the date of the first dispensing of drug is known, one can calculate, from the amount dispensed and the prescribed daily dose, the expected, or "legend," period of treatment. Thus, Dutch pharmacy records provide a sound basis for estimating the duration of drug exposure.⁴³

There are practical problems, however. The onset of drug exposure is putatively defined by the date of drug dispensing to the patient, but is subject to uncertainties about when the patient actually began taking the medicine, which is not necessarily coincident with the dispensing date. Greater uncertainty attends definition of the cessation of exposure. When there have been one or more refills of the same prescription, one can examine the intervals between refills for evidence of delay due to omitted doses, which is the most common form of noncompliance with prescribed drug regimens.⁴⁴ When the last refill has been dispensed, one faces the question of defining when

the last dose is likely to have been taken, and how long the effects of treatment may linger after the last dose was taken. An unfortunate lacuna in clinical pharmacology is its one-sided focus on studying, with great care and diligence, the temporal details of the onset of drug action, while almost totally ignoring the offset of drug action.⁴⁵ Most clinical pharmacologists assume that the time course of decline in the concentrations of drug and/or active metabolites dictate the time course of cessation of drug action, but that assumption is contradicted by too many examples of drug actions long outlasting any detectable drug in plasma.⁴⁵ Thus, the pharmacoepidemiologist who seeks to define the period of effective exposure is left having to make a number of assumptions. One that seems reasonable, pending experimental data, is to assume that the duration of drug action after the last dispensing is equal to 150% of the legend duration of the last prescription.³¹ However, if intervals between prior refills are appreciably longer than the legend durations, then one should frame a hypothesis consistent with refill intervals about the prevailing degree of poor compliance and adjust the estimates of total duration correspondingly.

This approach was applied by Petri *et al.* to assess a postulated association between the anti-migraine/antivertigo drug, flunarizine, and mental depression.³¹ That study, which pioneered the use of computerized Dutch pharmacy records, identified 1284 ever recipients of flunarizine, 180 of whom (14%) had, at one time or another during the study interval, been prescribed also an antidepressant. If, as postulated, flunarizine did trigger mental depression, one would have expected to see a flurry of antidepressant prescribing in the wake of flunarizine prescribing. Of course, in an ideal world, the appearance of signs of depression would lead physicians to terminate the prescription for flunarizine rather than add a prescription for an antidepressant, but the reality is that physicians are, on the whole, much more likely to start new prescriptions rather than to stop old ones, unless there is a very well characterized adverse effect of a previously prescribed drug. The data showed no clustering of antidepressant prescribing in the wake of flunarizine prescribing,

but did show clear evidence for channeling of flunarizine into use by patients who were unusually likely to be prescribed antidepressants, prior to or long after their use of flunarizine.

Using a subsequent prescription to treat an adverse effect was also applied by Sturkenboom *et al.*, who estimated the risk of vulvo-vaginal candidiasis among users of acitretin.⁴⁶ Prescriptions of drugs for treatment of vulvo-vaginal candidiasis were used as proxy for the occurrence of vulvo-vaginal candidiasis (sensitivity of 87%, specificity of 99%, respectively). The study showed a threefold-increased risk of vulvo-vaginal candidiasis during acitretin exposure.

DIMENSIONS OF DRUG EXPOSURE AND ITS CORRELATES

Important dimensions in interpreting the correlates of drug exposure are the patient's profile and health status. The profile includes age, gender, education, socio-economic status, and estimated compliance with prescribed pharmacotherapy. Health status is judged on the basis of information about disease severity and comorbidity—simple words that stand for considerable complexity. The difficulties in judging health status and prognosis are well illustrated by the analysis of Concato *et al.* of a purported increase in long term mortality following transurethral versus open prostatectomy.⁴⁷ An often overlooked factor whose importance cannot be overemphasized in judging health status is the patient's prescription drug history within the past several years.

Use of the prescription drug history alone as an indicator of health status and prognosis remains an underexplored topic. On its face, several things are evident. First, the writing of a prescription represents a summary medical judgment that the prescribed agent's indications are appropriate for the patient's present situation, as perceived and interpreted by the prescribing physician. Those perceptions and interpretations may be irrational or otherwise wrong, which one cannot, of course, judge from the prescription—though a clear clash between the indications and contra-indications of concomitantly prescribed agents may signal irrational prescribing. More information can, of

course, be gained from the physician's notes written around the same time, but at considerably greater cost. Whether the added information to be gleaned from physicians' notes justifies the "gleaning costs" depends on the importance of the question at issue. A second consideration is that drugs differ widely in the specificity of their indications, so the clinical information implicit in a first prescription will depend on the drug, its indications, and the other agents with which it is co-prescribed.^{17,48} Whether or not a first prescription is refilled, and for how long, gives an indication of whether it has struck enough of a perceived balance of good and ill to warrant its continuation. A third point is that many physicians prescribe certain drugs as a way of "buying time," to see after some days or weeks how a patient's complaints either resolve or become more indicative of a specific medical problem.

Obviously, the prescription drug history is far from a complete medical record, but can nevertheless tell a great deal. Several examples will illustrate. The finding of rifampicin, isoniazid, and ethambutol in a prescription drug history surely indicates the physician's conclusion that the patient has tuberculosis. The finding of a prescription drug history that, since 6 months ago, has included digoxin, captopril, furosemide, isosorbide dinitrate, sublingual nitroglycerin tablets, and procainamide, indicates the sudden appearance of coronary heart disease. The use of the prescription drug history is becoming also a useful tool to estimate the incidence and prevalence of disease, as shown in the example of epilepsy.⁴⁹

CHANNELING

Comparisons are the sinew of observational epidemiologic studies, but they depend on comparing like with like. One source for misclassification occurs when drugs with more or less identical pharmacology and indications are assumed to be used by patients with more or less identical profiles and health status.¹⁷ Figure 21.1(A), however, shows a prototypical description of the distribution of several patient populations along the dimension of disease severity. Patients included in randomized clinical trials (RCTs) often are positioned at the

right side of the figure in order to obtain enough impact versus placebo to show efficacy. In the example of Figure 21.1(A), usual recipients of drug A in everyday medical practice are positioned in the mild-moderate region. This shift in the severity of illness from the trials population to the general population of users probably reflects the change in pre- versus postmarketing uses of captopril, and is one of the reasons that the initially recommended dose was so much higher than that necessary for the vast majority of patients prescribed the drug in routine clinical practice.⁵⁰

Figure 21.1(B) shows rather an opposite picture of the RCT population, in the mild-moderate region of the population distribution of severity. This situation probably prevails in most RCTs, which exclude patients who have complex health problems associated with multiple diseases: the RCT population tends to be recruited from the low end of severity, in the interest of simplicity of analysis. Consequently, the profile and health status of patients prescribed drug B in routine practice may differ considerably from patients in the RCTs. Inman reported data supporting this situation in his PEM study of the osmotic pump formulation of indomethacin (Osmosin, Indosmos) and other NSAIDs⁵¹ (see also Chapter 14). The new formulation of indomethacin had been promoted as a form of the drug less likely than the conventional capsule form to produce gastrointestinal side effects, thereby prompting many physicians to choose the osmotic pump formulation for patients with a history of gastrointestinal problems. One of us has termed this phenomenon "channeling."⁵² It highlights one of the basic, often difficult-to-answer questions in pharmacoepidemiology: "did the drug bring the problems to the patient, or did the patient bring the problems to the drug?"¹⁷ One way to judge this matter is to compare the frequencies with which recipients of ostensibly identical agents are coprescribed other agents that indicate more severe disease. Some examples will be used illustrate this point.

Channeling may explain the flow of adverse reaction reports from an NSAID that was formulated and promoted in a manner very similar to that of Osmosin.^{51,53} In January 1989, SmithKline Beecham introduced into the Dutch

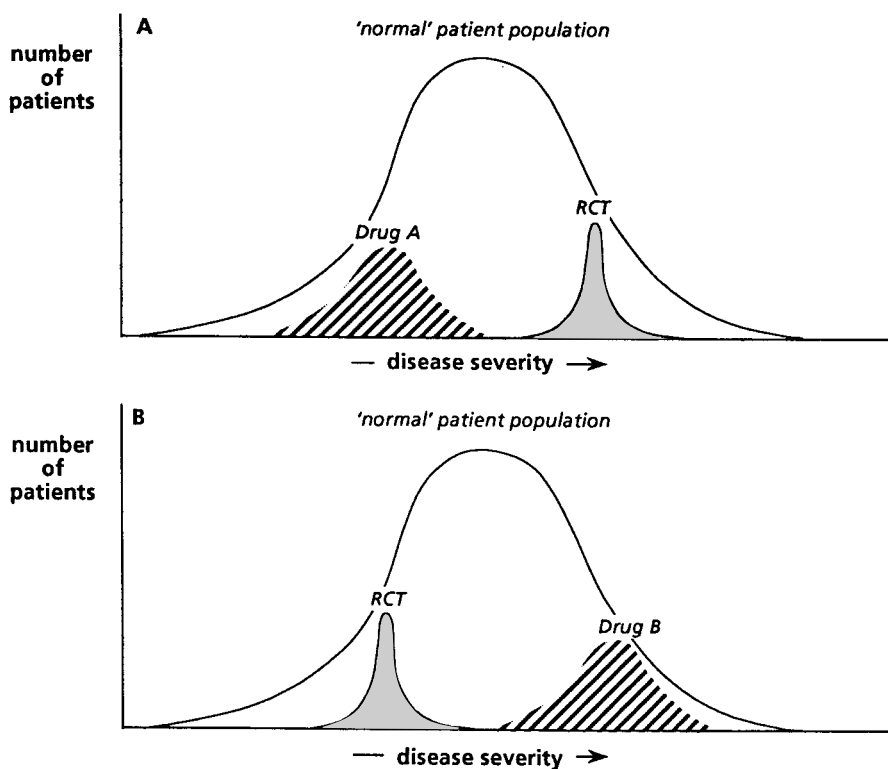


Figure 21.1. Two theoretical distributions of dosage severity in a "normal" patient population, in a randomized clinical trial population, and in patients receiving drugs in normal medical practice.

market a controlled release formulation (Oscorel, Oruvail) of the established NSAID ketoprofen. The product was heavily promoted with claims of safety with respect to gastrointestinal toxicity. In the first year after introduction of this product, the national authorities received a surge of case reports of gastrointestinal bleeding and/or perforation in recipients of Oscorel. This observation led to a pharmacy based study in which 837 patients receiving first prescriptions for Oscorel were identified. The drug use histories of these patients over the period 1987–1988 were searched for prescribing of agents indicated for the treatment of peptic ulcer disease. It was found that 24.1% of Oscorel recipients had received anti-ulcer drugs, versus 15.7% in a reference population (RR = 1.54; 95% CI 1.36–1.74). This pattern of channeling of Oscorel toward patients with a history of gastrointestinal disease was recognized in time to help interpret the adverse reaction

reports, and to prevent an incorrect conclusion, simply on the basis of adverse reaction reports, that the product was more prone than the conventional form of ketoprofen or other NSAIDs to cause gastrointestinal problems.

Another example where channeling has occurred is in the prescribing of the several inhalational β -2 agonists—salbutamol, terbutaline, and fenoterol. To evaluate the role of channeling in the controversy over the mortality risk attributed to fenoterol by Crane *et al.* in New Zealand, Petri *et al.* examined the extent to which fenoterol is channeled into use by high risk asthma patients, by studying the co-prescribing of β -2 agonists marketed in The Netherlands with systemic corticosteroids, as an indicator of asthma severity.⁵⁴ Compared to patients prescribed inhalational salbutamol, approximately twice the proportion of patients prescribed inhalational fenoterol were concomitantly prescribed systemic steroids;

patients prescribed inhalational terbutaline were intermediate with respect to the prescribing of systemic steroids. These data indicate that, in The Netherlands, fenoterol was channeled into use in patients with severe asthma. These different usage patterns appear to relate to the sequence with which the products were introduced into the market, with the later entering products having been promoted mainly to specialists for patients with more severe disease. Of course, the existence of channeling of a particular product in one country does not necessarily imply that the same channeling occurred in other countries, with the potential for different dates of marketing, different alternative therapies available, different medical practice, and different marketing activities.

Since the first channeling studies, other examples of substantial channeling have been found. Egberts *et al.* observed that patients receiving newly introduced antidepressants during the first year after their respective introductions were not comparable to patients who were receiving longer available antidepressants of the tricyclic class.⁵⁵ The recipients of the newly introduced agents tended to be patients who had failed to respond satisfactorily to older agents. Of course, failure to respond has a dual origin: pharmacologic nonresponse and clinically unrecognized noncompliance, so when the latter problem is the true reason for failed treatment, there is a high likelihood that the new treatment will similarly fail, unless the patient chooses that moment to commence proper dosing. It is important to recognize channeling, because it signifies that drugs that may be scarcely distinguishable on clinical pharmacological grounds nevertheless end up being used by patients with quite different health status.^{17,52} Careful analysis in attempting to adjust for differences in disease severity is needed to make unbiased comparisons among the groups of patients receiving one drug versus another.

UNDERSTANDING OF EXTRAORDINARY PROMOTIONAL MANEUVERS LEADING TO HYPERPRESCRIBING

Inman's PEM studies revealed that a small minority of physicians tend to generate disproportionately large fractions of prescriptions for newly

launched pharmaceuticals.⁵⁶ We have also observed this phenomenon in The Netherlands. It can take on remarkable proportions, with a half dozen physicians accounting for a third or more of all prescriptions country wide, with one or two physicians writing prescriptions numbering in the hundreds, which must have meant that virtually every patient who visited the doctor was given a prescription for the new agent. In some instances, the same physicians hyperprescribed one new product after another. We reported these findings in a letter to *The Lancet*, commenting that, in contrast to Inman's conclusions, the hyperprescribers had no distinguishing demographic or educational attributes.⁵⁷ Fortunately, our letter was published the day before headline news in *The Sunday Times* (London) reporting schemes by which some physicians received a substantial cash payment for each prescription for the new product. In following weeks, there ensued a series of such stories in the London papers involving several major pharmaceutical firms, each of which firmly denied the firm's involvement, attributing the reported maneuvers to rogue sales representatives and reporting dismissal of those responsible.

The seeding phenomenon takes many forms, of which the foregoing is an extreme, and the liberal distribution of micro-computers was perhaps somewhat less egregious. However distasteful these maneuvers may be, they are an expression of the intense economic pressures on new pharmaceuticals, which have incurred increasingly larger development costs that must be not only repaid and but also rewarded against current market standards for high risk investments, within the time period of the remaining life of applicable patents. Pharmacoepidemiologists, whatever their personal opinions about the ethical ramifications of these economic matters, are obliged to understand how these intense promotional efforts operate, for they are an integral part of the environment—just as, for example, microbiologists and parasitologists are not expected to “like” the various stratagems by which bacteria, viruses, and parasites infect or infest healthy humans, but it is their task to understand how these objects of their studies work. For this reason, we encourage students and researchers in pharmacoepidemiology

to undertake self-directed, *ad hoc* study of current promotional programs.

THE FUTURE

While the small size of The Netherlands has certain inherent disadvantages, it has a number of advantages, too. It is, for example, possible to have personal contact with all the leading experts in a given field, which creates the possibility of extensive "learning" before starting with "confirming."⁵⁸ Another feature that can facilitate pharmacoepidemiology research is the large number of villages in which a single pharmacy and five or six primary care physicians together constitute the core of pharmacotherapy in their area. This integration has recently gained official status in the Pharmacotherapy Discussion Groups, which, in their improvement of communication between pharmacists and physicians, have the potential to improve the quality of prescribing and the recognition and reporting of unexpected drug induced effects, beneficial as well as adverse.

The future will see more work on record linkage, to be sure, but also work to help define the contributions that the prescription drug history can make in the evaluation of health status, and the uses of objective, quantitative information on drug exposure and patient compliance to improve the quality of ambulatory health care.

With the introduction of new drug therapies in upcoming years, there will be a clear need for comparative risk/benefit evaluation after the drug has been used widely in daily practice. Risk/benefit comparisons of different drug therapies are subjects for observational epidemiologic studies, a logical next step after the conduct of randomized clinical trials designed to provide evidence for efficacy.^{19,29}

There is increasing evidence that bad response to drug therapy is correlated with both patient factors (disease severity, co-morbidity, metabolism, genetics) and utilization factors (quality of diagnosis, inappropriate prescribing, concomitant use of other drugs, noncompliance and other patterns of interrupted drug usage).^{58,59} However, comparisons in observational studies of the

risk/benefit balance of different drug therapies are often subject to scientific debate because factors such as confounding by indication, channeling bias, and other types of misclassification may cause flaws in the interpretation of the results of such studies.^{23,52} Because of commercial interests being an important feature of today's pharmaceutical marketplace, these questions need careful and independent scientific consideration. Identifying individual patient characteristics is considered an important strategy to improve like-with-like comparisons.⁴⁸

An important lead for future research is "patient typing," starting from the scope of epidemiology towards clinical, and molecular sciences as well. This approach requires the availability of extensive and longitudinal data on drug exposure, advanced medical record linkage systems, and an integrative knowledge base of pharmacoepidemiology, clinical epidemiology, and relevant fields in molecular epidemiology, e.g., genetics. There is increasing evidence that both the efficacy of drugs (e.g., cholesterol lowering agents) and their safety (e.g., Factor V Leiden and thrombotic risk) have their genetic correlates.^{7,35} These findings argue for innovative approaches to integrate research on drug exposure, patient characteristics of various types, and outcomes. Moreover, the impact of undertreatment is becoming increasingly visible and needs to be better understood.⁵⁹ The cry for evidence based pharmacotherapy should go hand in hand with this type of research, in order to provide the scientific basis for sound clinical decision making. In the interest of patient needs and public health in general, there is a great challenge ahead of us, to search for better methods and concepts for comparative risk/benefit evaluation in drug therapy.

REFERENCES

1. Kok FJ, Vandenbroucke JP, Van der Heide-Wessel C, Van der Heide RM. Dietary sodium, calcium, and potassium and blood pressure. *Am J Epidemiol* 1986; **123**: 134-8.
2. Jansen PAF, Schulte BPM, Meyboom RHB, Gribnau FWJ. Antihypertensive treatment as a

- possible cause of stroke in the elderly. *Age Ageing* 1986; **15**: 129–38.
3. van Ouweland FA, Gribnau FWJ. Nonsteroidal anti-inflammatory drugs as a prognostic factor in acute pulmonary edema. *Arch Intern Med* 1987; **147**: 176–9.
 4. 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.
 5. Vandenbroucke JP, Valkenburg HA, Boersma JW, Cats A, Festen JJM, Huber-Bruning O, Rasker JJ. Oral contraceptives and rheumatoid arthritis: further evidence for a preventive effect. *Lancet* 1982; **2**: 839–42.
 6. Hazes JMW, Dijkmans BAC, Vandenbroucke JP, Vries RPR, Cats A. Reduction of the risk of rheumatoid arthritis among women who take oral contraceptives. *Arthritis Rheum* 1990; **33**: 173–9.
 7. Bloemenkamp KW, Rosendaal FR, Helmerhorst FM, Buller HR, Vandenbroucke JP. Enhancement by factor V Leiden mutation of risk of deep-vein thrombosis associated with oral contraceptives containing a third-generation progestagen. *Lancet* 1995; **346**: 1593–6.
 8. Helmerhorst FM, Bloemenkamp KW, Rosendaal FR, Vandenbroucke JP. Oral contraceptives and thrombotic disease: risk of venous thromboembolism. *Thromb Haemostas* 1997; **78**: 327–33.
 9. Esdaile JM, Horwitz RI. Observational studies of cause–effect relationships: an analysis of methodologic problems as illustrated by the conflicting data for the role of oral contraceptives in the etiology of rheumatoid arthritis. *J Chron Dis* 1986; **39**: 841–52.
 10. Meyboom RHB, Ferrari MD, Dieleman BP. Parkinsonism, tardive dyskinesia, akathisia and depression induced by flunarizine. *Lancet* 1986; **2**: 292.
 11. Stricker BHCh, Groot RRM, Wilson JHP. Glafenine-associated anaphylaxis as a cause of hospital admission in the Netherlands. *Eur J Clin Pharmacol* 1991; **40**: 367–71.
 12. Stricker BHCh, Tijssen JGP. Serum sickness-like reactions to cefaclor. *J Clin Epidemiol* 1992; **45**: 1177–84.
 13. Egberts AC, Meyboom RH, De Koning GH, Bakker A, Leufkens HG. Non-puerperal lactation associated with antidepressant drug use. *Br J Clin Pharmacol* 1997; **44**: 277–81.
 14. van der Klauw MM, Wilson JH, Stricker BH. Drug-associated agranulocytosis: 20 years of reporting in The Netherlands (1974–1994). *Am J Hematol* 1998; **57**: 206–11.
 15. Haayer-Ruskamp FM. Drug utilization studies in The Netherlands. *Pharm Weekbl Sci* 1990; **12**: 91–6.
 16. Hekster YA, Vree TB. Drug utilization research in clinical practice. *Drug Intell Clin Pharm* 1986; **20**: 679–82.
 17. Leufkens HG, Urquhart J. Variability in patterns of drug usage. *J Pharm Pharmacol* 1994; **46**: 433–7.
 18. Ellwood P. Shattuck lecture: outcomes management. *New Engl J Med* 1988; **318**: 1549–56.
 19. Psaty BM, Siscovick DS, Weiss NS, Koepsell TD, Rosendaal FR, Lin D, Heckbert SR, Wagner EH, Furberg CD. Hypertension and outcomes research. From clinical trials to clinical epidemiology. *Am J Hypertens* 1996; **9**: 178–83.
 20. Manten B, Westendorp RG, Koster T, Reitsma PH, Rosendaal FR. Risk factor profiles in patients with different clinical manifestations of venous thromboembolism: a focus on the factor V Leiden mutation. *Thromb Haemostas* 1996; **76**: 510–3.
 21. Schrijvers AJP, ed. *Health and Health Care in The Netherlands. A Critical Self-Assessment by Dutch Experts in the Medical and Health Sciences*. Utrecht: De Tijdstroom, 1998.
 22. Leufkens HG, Haaijer-Ruskamp F, Bakker A. G Dukes. Scenario analysis of the future of medicines. *Br Med J* 1994; **309**: 1137–40.
 23. 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.
 24. Visser LE, Stricker BH, van der Velden J, Paes AH, Bakker A. Angiotensin converting enzyme inhibitor associated cough: a population-based case-control study. *J Clin Epidemiol* 1995; **48**: 851–7.
 25. Van der Linden PD, Van der Lei J, Vlug AE, Stricker BHCh. Skin reactions to antibacterial agents in general practice. *J Clin Epidemiol* 1998; **51**: 703–8.
 26. 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.
 27. Borden EK, Lee JG. A methodologic study of postmarketing drug evaluation using a pharmacy-based approach. *J Chron Dis* 1982; **35**: 803.
 28. Stricker BHCh, Barendregt M, Herings RMC, De Jong-van den Berg LTW, Cornel MC, De Smet PAGM. Ad hoc tracing of a cohort of patients exposed to acitretine (Neotigason) on a nation wide scale. *Eur J Clin Pharmacol* 1992; **42**: 555–7.
 29. Urquhart J. Time to take our medicines, seriously (inaugural professorial lecture, University of Limburg, April 3, 1992). *Pharm Weekbl* 1992; **127**: 769–76.
 30. Leufkens HG, Ruter EM, Ameling CB, Hekster YA, Bakker A. Linkage of pharmacy data on heavy users of nonsteroidal anti-inflammatory drugs to information from general practitioners. *J Pharmacoepidemiol* 1991; **2**: 67–77.

31. 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.
32. Herings RMC, Stricker BHC, Nap G, Bakker A. Pharmaco-morbidity linkage: a feasibility study comparing morbidity in two pharmacy-based exposure cohorts. *J Epidemiol Community Health* 1992; **46**: 136–40.
33. Herings RM, Stricker BH, de Boer A, Bakker A, Sturmans F. Benzodiazepines and the risk of falling leading to femur fractures. Dosage more important than elimination half-life. *Arch Intern Med* 1995; **155**: 1801–7.
34. Herings RM, de Boer A, Stricker BH, Leufkens HG, Porsius A. Hypoglycaemia associated with use of inhibitors of angiotensin converting enzyme. *Lancet* 1995; **345**: 1195–8.
35. Kuivenhoven JA, Jukema JW, Zwinderman AH, de Knijff P, McPherson R, Brusckhe AV, Lie KI, Kastelein JJ. The role of a common variant of the cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. The Regression Growth Evaluation Statin Study Group. *New Engl J Med* 1998; **338**: 86–93.
36. Straus SMJM, Meiners AP. Postmarketing surveillance in the pharmacy. *Pharm World Sci* 1998; 48–50.
37. Vries CS de, Diepen NM van, Tromp ThFJ, de Jong van den Berg LTW. Auditing GP's prescribing habits: medication initiated by specialists. *Eur J Clin Pharmacol* 1996; **50**: 349–52.
38. de Jong van den Berg LTW, van den Berg PB, Haayer-Ruskamp FM, Duker MNG, Wesseling H. Investigating drug use in pregnancy. Methodological problems and perspectives. *Pharm Weekbl Sci* 1991; **13**: 32–8.
39. de Jong PCM, Huijsmans AA, Nienhuis HE, Nijdam WS, Zielhuis GA, Eskes TKAB. Validation of a questionnaire on medical drug use during pregnancy. *Am J Epidemiol* 1992; **134**: 998–1002.
40. Lau HS, de Boer A, Beuning KS, Porsius A. Validation of pharmacy records in drug exposure assessment. *J Clin Epidemiol* 1997; **50**: 619–25.
41. Milgrom H, Bender B, Ackerson L, Bowry P, Smith B, Rand C. Noncompliance and treatment failure in children with asthma. *J Allergy Clin Immunol* 1996; **98**: 1051–7.
42. B. Mengden T, Medina RMH, Beltran B, Alvarez E, Kraft K, Vetter H. Reliability of reporting self-measured blood pressure values by hypertensive patients. *Am J Hypertens* 1998; **11**: 1413–7.
43. Staa TP van, L Abenheim L, Leufkens HG. A study of the effects of exposure misclassification due to the time-window design in pharmacoepidemiologic studies. *J Clin Epidemiol* 1994; **47**: 183–9.
44. Urquhart J. The electronic medication event monitor—lessons for pharmacotherapy. *Clin Pharmacokinetics* 1997; **32**: 345–56.
45. Urquhart J. Pharmacodynamics of variable patient compliance: implications for pharmaceutical value. *Adv Drug Delivery Rev* 1998; **33**: 207–19.
46. Sturkenboom MC, Middelbeek A, de Jong van den Berg LT, van den Berg PB, Stricker BH, Wesseling H. Vulvo-vaginal candidiasis associated with acitretin. *J Clin Epidemiol* 1995; **48**: 991–7.
47. Concato J, Horwitz RI, Feinstein AR, Elmore JG, Schiff SF. Problems of comorbidity in mortality after prostatectomy. *J Am Med Assoc* 1992; **267**: 1077–82.
48. Grobbee DE, Hoes AW. Confounding and indication for treatment in evaluation of drug treatment for hypertension. *Br Med J* 1997; **315**: 1151–4.
49. Shackleton DP, Westendorp, RGJ, Kasteleijn-Nolst Trenité DGA, Boer A de, Herings RMC. Epileptic medication: a road to determining the number of individuals with seizures. *J Clin Epidemiol* 1997; **50**: 1061–8.
50. Temple R. Dose-response and registration of new drugs. In: Lasagna L, Erill S, Naranjo CA, eds, *Dose-Response Relationships in Clinical Pharmacology*. Amsterdam: Elsevier, 1989; 145–67.
51. Inman B. Comparative study of five NSAID's. *PEM News* 1985; **3**: 3–134.
52. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med* 1991; **10**: 577–81.
53. Leufkens HG, Urquhart J, Stricker BHCh, Bakker A, Petri H. Channeling of controlled release formulation of ketoprofen (OSCOREL) in patients with history of gastrointestinal problems. *J Epidemiol Community Health* 1992; **46**: 428–32.
54. Petri H, Urquhart J, Herings R, Bakker A. Characteristics of patients prescribed three different inhalational beta-2 agonists: an example of the channeling phenomenon. *Postmarket Surveill* 1991; **5**: 57–65.
55. Egberts ACG, Lenderink AW, Koning de FHP, Leufkens HGM. Channeling of three newly introduced antidepressants to patients not responding satisfactorily to previous treatment. *J Clin Psychopharmacol* 1997; **17**: 149–55.
56. Inman W, Pearce G. Prescriber profile and post-marketing surveillance. *Lancet* 1993; **342**: 658–61.
57. Leufkens HG, Urquhart J. Prescriber profile and postmarketing surveillance [letter]. *Lancet* 1993; **342**: 1178.
58. Sheiner LB. Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* 1997; **61**: 275–91.
59. Klungel OH, Boer A de, Paes AHP, Seidell JC, Nagelkerke NJD, Bakker A. Undertreatment of hypertension in a population-based study in The Netherlands. *J Hypertens* 1998; **16**: 1371–78.

22

The Tayside Medicines Monitoring Unit (MEMO)

JOSIE M.M. EVANS

Medicines Monitoring Unit, Department of Clinical Pharmacology, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK

THOMAS M. MACDONALD

Medicines Monitoring Unit, Department of Clinical Pharmacology, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK

INTRODUCTION

The Medicines Monitoring Unit (MEMO) is a university based organization that was set up to carry out studies to detect and quantify serious drug toxicity in the community, using record linkage techniques. MEMO has access to data generated within the United Kingdom National Health Service (NHS). This service is tax funded and free at the point of consumption. There are thus no socioeconomic eligibility distinctions and the level of health care given to an individual is based on need alone. Currently MEMO utilizes 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 with, for example, only 5% of nearly 4000 cimetidine takers lost to follow-up in a five year period.¹

Health care for the region is coordinated by the Tayside Health Board, which maintains a computerized record of all patients registered with general practitioners as well as data on inpatient hospital morbidity and mortality. These data are available to MEMO and form the backbone of the record linkage system.

The potential of the MEMO system for pharmacoepidemiology research was described in 1994.² Since then, this potential has clearly been realized, as descriptions of completed and current research in this Chapter will show. Although MEMO was originally set up for pharmacoepidemiology research and this is still the main focus of its research activities, recent advances mean that studies in outcomes research, general epidemiology, and economics are also possible.

DESCRIPTION

PATIENT IDENTIFICATION

Every person who is 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, the ninth digit indicating sex, and the tenth digit incorporating a check sum to ensure the validity of the number. 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 the date of death. The demographic breakdown of the Tayside population can therefore be readily obtained.

The CHI number is used as the patient identifier for all health care activity in Tayside, be this primary care or hospital inpatient care. The number can be found by combining the name with the date of birth and performing a computer search. This unique number allows efficient linkage of records of patient activity and outcome, and once it is allocated it never changes.

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 GP prescribing. When a patient receives a prescription from his doctor, he takes it to any of the community pharmacies where the prescription is dispensed. The pharmacist then sends the original prescription form to the Pharmacy Practice Division of the Common Services Agency (PPD) in Edinburgh to obtain reimbursement (Figure 22.1).

After paying the pharmacists and dealing with any appeals, the PPD sends the cashed prescription forms to MEMO, where they are stored on a database linked to the CHI number. The following system has been developed to perform this task.

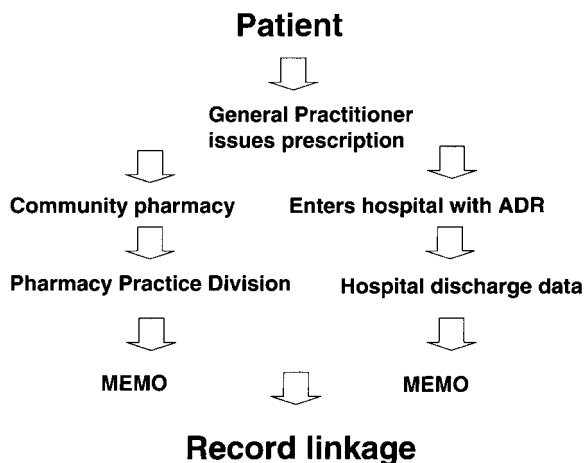


Figure 22.1. An outline of the data flow in record linkage drug safety studies.

ASCRIBING THE CHI NUMBER TO PRESCRIPTIONS

All prescriptions contain the patient's name and address (sometimes handwritten and barely legible), but not the date of birth. In addition, the prescription lists the prescribed drug name (generic or proprietary), the dispensed drug (generic or proprietary), the drug dose, the total number of doses, the dosing instructions, and the GP code number. In order to handle this information efficiently, a custom made, interactive, menu driven computer system has been developed. Using this system, it is possible to determine the CHI number from the patient details on the prescription, even if only some of these details are readable. The program regards each part of the name and address as a potentially unique patient-identifying symbol. The entire Health Board Master Patient Index is indexed by these symbols. For any patient, one or more of these symbols is all that is required for identification. Thus, if a patient lives at 269 Perth Road, and is called MacDonald, simply entering "MacDonald" and "269" is usually sufficient for identification. Alternatively, typing in "269" and "Perth" lists all those occupants of addresses in Tayside which contain the terms 269 and Perth. A unique GP code appears at the bottom of each prescription

form and is also a useful symbol. Each person is registered with a GP and so should appear on that GP's list only. To cope with patients seeing other GPs in the same practice, the computer checks that the GP code on the prescription belongs to a GP in the practice.

This system ascribes in excess of 98% of CHI numbers to prescriptions on the first pass. The remaining 2% consist of persons who have recently changed address within Tayside or moved into Tayside from outside, and those visitors that are not registered with GPs. The Community Health Master Patient Index is continually updated by the Health Board, but nevertheless the copy held in MEMO is always a little out of date. Consequently, data on those patients not identified on the first pass are entered manually and reconciled when the CHI number update data are available.

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 course of treatment.

DRUG DATA

The date the prescription was written is recorded and then the individual drug code is entered using the PPD's own coding system. This has some limitations in that it is simply a list of all products currently reimbursed. Generic and proprietary names are not linked and neither are pack sizes. Furthermore, because it is simply a drug list, constituent parts of products (for example triamterene, which forms the potassium sparing element of many diuretic combinations) cannot be automatically linked and this process has to be carried out manually. However, there are several more sophisticated drug dictionaries now available and it is likely that one may be adopted by MEMO in the future.

The data clerk entering prescribing data selects the drug preparation from a menu of all available products. This is important as it avoids miscoding of preparations and spelling errors. Where a drug is prescribed generically but dispensed in proprietary form (e.g., sudenaphil is only available as Viagra) this is recorded. The total amount of drug dispensed is also entered together with the dosing instructions, thus allowing the duration of any prescription to be calculated (Table 22.1). To ensure accuracy, regular quality control checks are carried out by dual entry of a proportion of prescription items.

MEMO currently employs clerks to enter these data and, although their collection is relatively labor intensive and expensive, in its favor is that it is under the direct control of MEMO and that the data are held in a format suitable for the needs of record linkage. Since January 1993, data on all drugs dispensed in Tayside have been collected, although prior to this data were collected on specific drugs only (notably non-steroidal anti-inflammatory drugs, ulcer healing drugs and hormone replacement therapy). MEMO now has records of over 15 million prescriptions dispensed in Tayside between 1989 and 1996.

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. In Tayside this is coded by medical clerks before being entered onto computer and subjected to quality control. The data are then

Table 22.1. Prescribing data

Drug name prescribed (generic or proprietary)
Drug name dispensed (generic or proprietary)
Dose
Form
Number of doses issued
Instructions
Date
Prescribing doctor

sent to the Information and Statistics Division (ISD) of the Common Services Agency of the National Health Service. Each SMR1 record has one principal and five other diagnostic fields coded according to the International Classification of Diseases ninth revision (ICD9).³ There is also one main operation or procedure field and three others coded according to the Office of Population and Census Surveys fourth revision (OPCS4) classification.⁴ In 1996, the NHS introduced the tenth revision of the ICD codes,⁵ and there is still debate as to whether Read codes will be introduced.⁶ In Tayside there are approximately 63 000 hospital discharges per year. MEMO has in-house historical SMR1 data going back to 1980. These data allow for a past medical history of hospitalization for a condition to be controlled.

The SMR1 database contains details of deaths certified in hospital, which may be up to 85% of the total mortality.¹ 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 dataset that is indexed by the CHI number can be linked into MEMO's record-linkage database, including other Scottish Morbidity Record returns supplied by ISD. In MEMO, commonly used datasets are the cancer registration database (SMR6), child development records, maternity records (SMR2), psychiatric records (SMR4), and neonatal discharges (SMR11).

OTHER OUTCOME DATASETS

Some health care datasets are available in Tayside that are not indexed by CHI number. However, provided that some patient demographic detail is present, such as name, date of birth, and postcode, 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, subjects involved in 22 000 road traffic accidents in Tayside.

VALIDATION OF HOSPITAL DATA

One of the strengths of MEMO is that the original case records of hospitalized patients in Tayside can be examined. This allows for quality control of the computerized data and can also deal with some elements of confounding. For example, persons admitted with gastrointestinal bleeding are nearly always asked about their previous ingestion of aspirin and nonsteroidal anti-inflammatory drugs. Over-the-counter (OTC) preparations are not recorded on 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.⁸

PRIMARY CARE DATA

Progressively more GPs are using computerized systems to aid patient management, although at present they are not available to MEMO for record linkage. However, it is possible to abstract written records in primary care manually and research nurses in MEMO have been granted access to primary care records for specific studies.^{7,8}

OTHER INFORMATION

Since all patients and their addresses are known, including post code, and information is available from the decennial census regarding the relative deprivation levels of postcode areas, the so-called Carstairs deprivation score can be used as a relatively crude indicator of the socioeconomic status of patients.^{9,10}

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 Tayside Health Board, and uses this to track the population of patients alive and resident in Tayside to define study populations for drug safety studies. 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 that have been dispensed at a pharmacy and so primary noncompliance is eliminated. In a

study carried out to assess the extent of primary noncompliance in Tayside, a large family practice (11 500 patients) wrote all prescriptions in duplicate (carbon copy) forms over a three month period.¹¹ The copies were sent to MEMO. The original top copy forms which were redeemed by patients at community pharmacies were also returned to MEMO by the PPD. Duplicate forms for which no original was present represented the prescriptions which were not redeemed. Figure 22.2 shows the rate of primary noncompliance by age and sex, and it is clear that in some age groups this is a significant problem.

Prescribing may be influenced by the experience and interests of particular GPs, and this might distort studies in certain conditions. For example, Figure 22.3 shows the prescribing of misoprostol by GPs in Tayside between January 1989 and December 1991. There are clear variations in prescribing, with the majority of the prescribing being done by the minority of GPs. Since the list of patients registered with individual GPs and groups of GPs (practices) are known and the GP who initiated the therapy is recorded, it is possible to pick comparator groups from patients under the care of the same (or different) physician(s) or practice(s). Thus, it is possible to control for confounding by prescriber behavior.

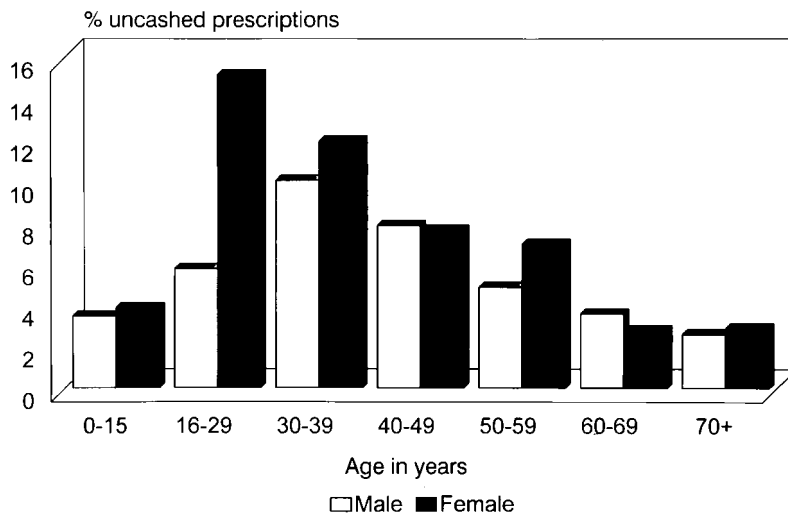


Figure 22.2. The percentage of uncashed prescriptions by age and sex (primary noncompliance).

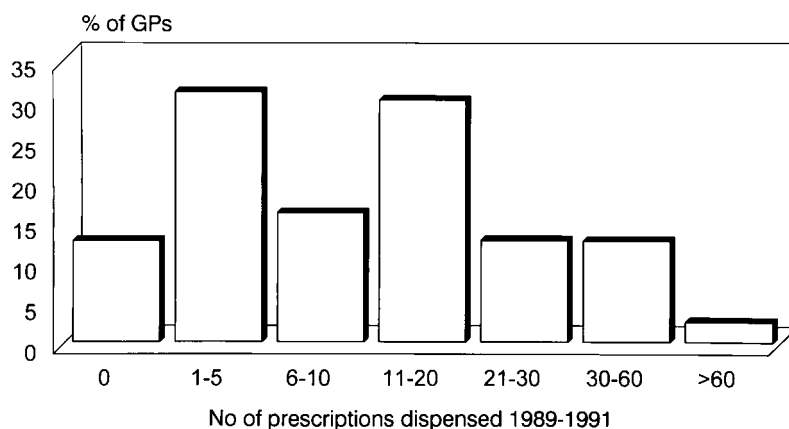


Figure 22.3. The percentages of general practitioners prescribing various amounts of misoprostol in Tayside between January 1989 and December 1991, demonstrating that some never prescribe this drug, whereas others prescribe it frequently.

A further advantage in Tayside is that there is currently no structure to inhibit the prescribing of newly marketed drugs. Thus, studies of new agents with high market penetration are possible.

ACCESSIBILITY TO MEDICAL RECORDS

A major strength of MEMO is the ability to examine original hospital case records where necessary. Several studies validating the computerized diagnostic data with the case records have been carried out, with variable results (see next section) depending on the criteria used. Within the NHS such case record searching for the purposes of drug safety evaluation is ethically permissible.¹²

WEAKNESSES

The current population of Tayside is approximately 400 000 and is comparatively small for pharmacoepidemiology studies, even for the study of relatively commonly prescribed drugs. In addition, 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). Scottish doctors are conservative prescribers of new drugs, so the market penetration of new agents tends to be low in

the first few years after launch. This limits the ability to study new chemical entities, arguably the most important drug group to study. Offsetting these disadvantages, the profile of certain diseases, for example cardiovascular disease, is higher in Scotland than in other populations and consequently the prescribing of drugs used in the prevention and treatment of these diseases is proportionately higher.

Another weakness, but one that is common to most drug safety databases, is the inability to capture directly exposure to OTC drugs or drugs prescribed in hospital. Perhaps more importantly, given that confounding by indication is arguably one of the most difficult potential sources of error in pharmacoepidemiology research, the diagnostic indication for prescribing is not available. Where the indication for a drug is wide (e.g., beta adrenoceptor blocking drugs can be given for indications varying from anxiety to hypertrophic cardiomyopathy), this can lead to considerable difficulties.

MEMO has undertaken not to contact patients directly to elicit information on possible confounding factors without the explicit permission and cooperation of GPs. However, primary care and hospital records can be checked and, although the quality varies, some data on smoking and alcohol can be retrieved from them.⁸ This is also a method

by which to discern outpatient diagnoses, which are not available electronically for record linkage in MEMO. MEMO is therefore currently best suited for the study of serious drug toxicity that requires hospital admission.

One of the criticisms leveled at record linkage studies is the inaccuracy of computerized medical diagnoses. The discharge diagnoses for SMR1 are abstracted from the clinical discharge summaries by specially trained coding clerks. These clerks on occasions have to interpret "soft diagnoses," such as symptoms for which no cause can be found. In addition, non-standard terminology may be employed to describe an illness, for example eponymous terms, and 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. Several validation studies of the accuracy of hospital discharge data in Scotland have been performed, comparing the coded diagnoses with diagnoses inferred by one or more senior doctors who have reviewed the original case records.¹³⁻¹⁵ The most pertinent of those studies carried out on Tayside data found 18% of internal medicine diagnoses to be clinically unacceptable.¹³ Since the publication of this study, steps have been taken, mainly for resource management reasons, to improve the diagnostic accuracy of computerized data by involving clinicians in quality control. This initiative has improved substantially the diagnostic accuracy of records.

In MEMO, the coded diagnoses are usually checked individually for drug safety studies. This is a labor intensive process. It is also arguable whether nondifferential misclassification is a major problem, or whether the selection bias that could be introduced if not all records are available for checking is potentially more damaging.¹⁶

Another problem 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 under the care of a physician with a gastrointestinal hemorrhage and is transferred to the care of a surgeon, transferred to an intensive care specialist postoperatively, and then back to a physician prior to 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 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.

PARTICULAR APPLICATIONS

A considerable amount of development work has gone into MEMO over several years and a suite of custom written software is available to perform record linkage drug utilization studies, case-control studies, and cohort studies. Prior to 1990, several studies showed the potential viability of the record linkage system,^{1,17,18} but these studies were tedious to perform, since the data were distributed between Health Board and university mainframe computers and much of the analysis had to be performed manually. The software developed by MEMO allows interactive searching of the databases by non-computer-literate researchers, thus allowing a rapid response when necessary.

DRUG UTILIZATION RESEARCH

MEMO is able to produce very detailed drug utilization data, broken down by age, sex, date, day of the week prescribed, prescriber, generic or proprietary dispensing, co-prescribing, acute prescribing and/or repeat prescribing, dose, and duration (Figure 22.4). Specific studies can then be carried out to examine in more detail how drugs are prescribed and used in Tayside, from the standpoint of either the prescriber or the patient. This is important as the safety and efficacy of drugs may depend on whether they are used properly.

One important dimension is the audit of GP prescribing in the population, although GP specific data are analyzed anonymously and individual

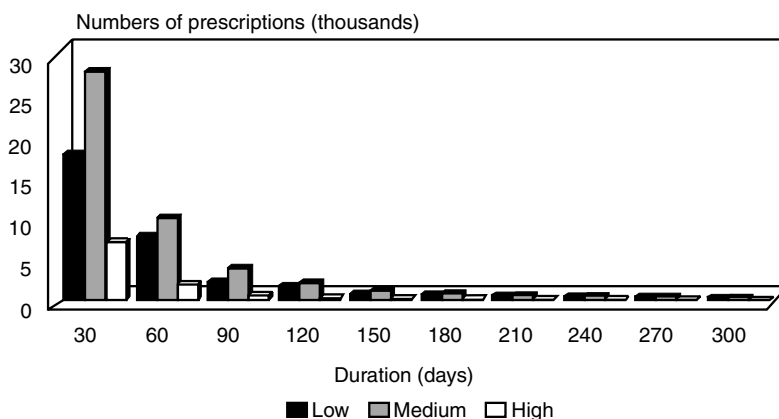


Figure 22.4. Duration of new prescribing of nonsteroidal anti-inflammatory drugs (NSAIDs) by dose.

GPs are never identified. For example, one study identified rare instances of potentially hazardous co-prescribing of β -antagonists and β -agonists to patients in Tayside likely to have asthma or chronic obstructive airways disease, by linking the dispensed prescribing database to hospital admission records.¹⁹ Another study exploited the detailed data on dose and duration (inferred from the total amount dispensed and the dosage instructions) and showed that various anti-depressants are often prescribed at ineffective doses for insufficient durations.²⁰ For example, Figure 22.5 shows that the majority of patients prescribed tricyclic antidepressants (TCAs) receive subtherapeutic doses, although this is not the case for the newer selective serotonin re-uptake inhibitors

(SSRIs). The processing of prescribing data according to the demographic characteristics of prescribing GPs has also yielded some useful insights into the characteristics of “good” prescribers. For example, differences in the prescribing of antibiotics and psychotropic medication have been seen between GP registrar training and non-training practices.²¹

Prescribing might also vary by patient factors, independent of need or disease severity, an example being the variation in the use of hormone replacement therapy by socioeconomic status.¹⁰ A related issue is patient compliance with medication, and by assessing how medication is collected by patients, in terms of numbers of prescriptions dispensed and intervals between them, and linking

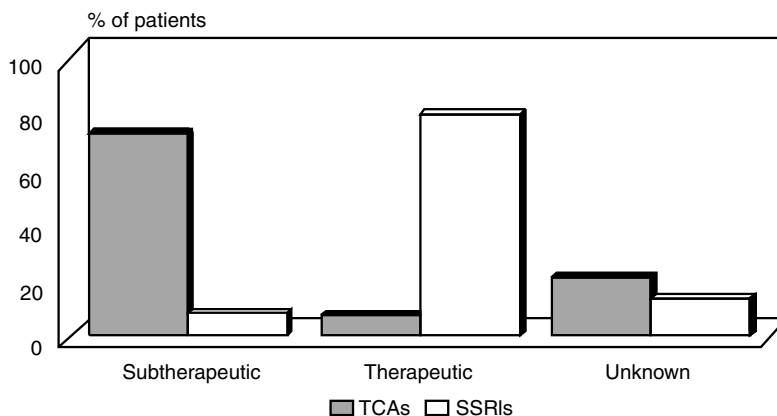


Figure 22.5. The percentages of patients prescribed antidepressant drugs at therapeutic and subtherapeutic doses.

this to outcome datasets, it is possible to assess the effects of noncompliance. For example, a study in diabetes showed that adolescents in Tayside who have “brittle” diabetes are often noncompliant with insulin.²²

DRUG SAFETY RESEARCH

While development was under way, software was tested by running several studies using a database of primary care prescribing in a practice of 11 000 patients dating from 1985. As an example, Figure 22.6 shows a graph illustrating the power of a simple plotting program to detect the occurrence of vaginal candidiasis following the ingestion of oral antibiotics in women.²³ The occurrence of a prescription for a vaginal antifungal drug was taken as a surrogate measure of candidiasis and the occurrence of this prior to and following any oral antibiotic was plotted. It is clear that there is an excess of candidiasis within 30 days of commencing antibiotics. In order to confirm this relationship, a case–control study was mounted, with the risk period varied sequentially by six days. Thus, the study was run examining antibiotic prescribing in the six days prior to a prescription for an antifungal agent, then it was repeated for the period between 7–12 days prior, then 13–18

days, etc. The results broadly confirmed the previous finding, demonstrating a similar pattern of risk with the overall risk for the 28 day period being 5.5 (95% CI 3.8 and 7.9).

Numerous drug safety studies have been completed in MEMO since this early work. For example, cohort study designs have been used to evaluate the risk profile of nonsteroidal anti-inflammatory drugs (NSAIDs). Although the increased risk of upper gastrointestinal complications associated with NSAID use is well established,²⁴ the large number of study subjects and the additional information available in MEMO has allowed investigation that is more detailed. For example, a cohort study among 78 191 patients newly exposed to NSAIDs and 78 207 unexposed comparators, showed that there was an increased risk only among patients without a history of upper gastrointestinal events.²⁵ Another study in 50 000 subjects investigated the risk with duration of use, and found that it remained constant with continuous exposure,²⁶ in contrast to previous findings.²⁷

The case–control method is a study design often chosen for its efficiency, as the number of study subjects required is usually less than that of a cohort study. In MEMO, when a study involves checking the original medical records of patients to

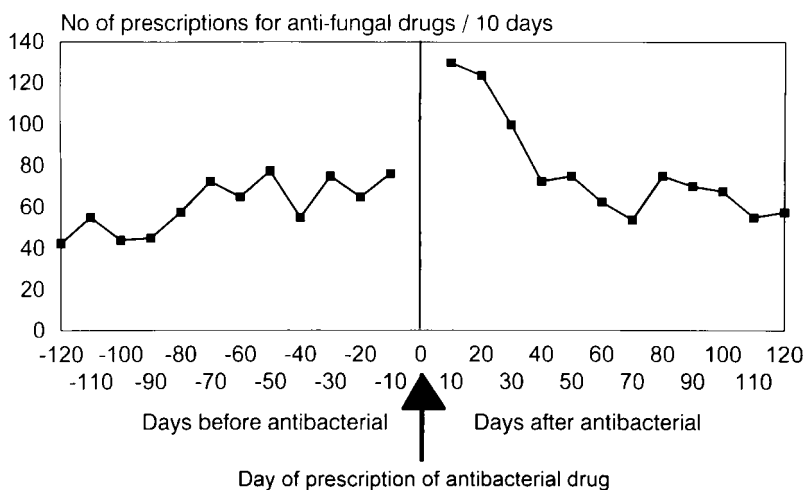


Figure 22.6. A plot of the prescribing of vaginal antifungal drugs both before and after the prescribing of an antibacterial drug in women. Note the clear excess of prescribing of antifungal drugs in the 30 days following the prescription of an antibacterial drug.

validate diagnoses or to obtain additional information on confounding factors, this is often an important consideration. For case-control studies in MEMO, study populations are rigorously defined, for example only including patients who were resident in Tayside during the entire study period (to ensure complete exposure histories are available). Cases are then identified from one of the SMR databases, and matched sets of either "community" controls (from the study population) or "hospital" controls (from patients admitted to the same hospital as the case but with a diagnosis unrelated to exposure) can be generated. Experience has shown that risks are more likely to be overestimated using community controls, and underestimated using hospital controls, therefore the nature of the risk may be important here. For example, arguably it is preferable to overestimate, rather than underestimate, the risk of a lethal side effect. Exposure frequencies can be calculated for cases and controls using data from the dispensed prescribing database, and exposure odds ratios calculated. It is also possible to adjust for confounding from other drug use.

The case-control design has been used in a range of MEMO studies investigating the side-effect profile of topical and oral NSAIDs. These studies have shown, for example, that oral NSAIDs, but not topical NSAIDs, are implicated in hospitalization for upper gastrointestinal hemorrhage and perforation,²⁸ acute renal failure,²⁹ and acute colitis,³⁰ but that they are unlikely to be associated with acute appendicitis.³¹

The case-crossover design was employed in a study examining the risks of road traffic accidents associated with benzodiazepine use.³² This design is suitable for the evaluation of transient risks, and because cases are used as their own controls, problems of confounding can be dealt with neatly.

METHODOLOGICAL STUDIES

MEMO has also provided the setting for some methodological work in pharmacoepidemiology. For example, an investigation into the selection of study subjects and the relative importance of misclassification and selection bias was completed for the case-control method,¹⁶ in addition to a

comparison of different data sources for information on confounding "lifestyle" factors, such as smoking and alcohol.⁸ Work on the cohort method has focused on the selection of comparator groups and whether the length of "risk windows" influences the estimate of effect.³³ This is the length of time after a drug has been prescribed during which a potential side-effect might be identified. Work on sample sizes for cohort studies has also been published.³⁴

OTHER RESEARCH APPLICATIONS

An interesting application of record-linkage in MEMO is the compilation of patient-specific disease registers using information from different sources. For example, a validated diabetic register of all patients (treated or nontreated) with type 1 and type 2 diabetes has been constructed.⁷ This is known as the DARTS initiative (Diabetes Audit and Research in Tayside Scotland), a collaboration between MEMO, all GP practices in Tayside and the Diabetes Units in three Health Care Trusts. The register is compiled by record-linking data from eight independent data sources: Scottish Morbidity Record 1 (SMR1), four diabetes clinics in three Health Care trusts, and a mobile eye van that has been operating in Tayside since 1990 performing community retinopathy screening, all of which use the CHI number routinely as a patient identifier. Data from MEMO's dispensed prescribing database (which includes glucose monitoring equipment) and the results of biochemistry tests in hospitals are also used. The register has been used for pharmacoepidemiology research.^{22, 35}

The incorporation of cost into the MEMO dispensed prescribing database also makes this an important resource for pharmaco-economic research, particularly that which recognizes the cost of adverse reactions to drugs that occur in the community,³⁶ an aspect often overlooked. There are also other more general health economics opportunities. One recent study assessed the cost-effectiveness of antibiotic treatment in a large health care center in Tayside.³⁷ An interesting evaluation of the economic consequences of hip replacement showed that while this procedure

reduced the costs of both prescribing for NSAIDs and ulcer-healing drugs, and also hospitalization for upper gastrointestinal hemorrhage, there were more likely to be hospital admissions for other conditions.³⁸

THE FUTURE

Dispensed prescribing data collection in MEMO is labor intensive and expensive. The automated capture of computerized dispensed prescribing data has therefore been investigated in five test pharmacies, a method which could eventually become Tayside-wide or Scottish-wide.³⁹ Client-server systems have been developed which use Integrated Services Digital Network (ISDN) links between community pharmacies and MEMO. MEMO makes available to pharmacy computer systems the CHI database along ISDN lines, and the CHI number can be allocated to data collected by the pharmacy at source. These data are then transmitted to MEMO immediately. Of 200 000 prescription items for which data were collected using this methodology, a comparison with a sample of duplicate data collected by MEMO in the usual way showed that there was agreement for 98% of the items.

As previously discussed, the underlying population of Tayside used for MEMO's drug safety studies is too small for accurate quantification of the risks of adverse drug reactions that occur at low frequency in the population (either because the drug is rarely prescribed or because the reaction itself is rare). A project has therefore been carried out to test the feasibility of performing record-linkage projects in the entire population of Scotland (5.5 million).⁴⁰ This is a collaboration between MEMO, the Pharmacy Practice Division (PPD), and the NHS Information and Services Division (ISD). Three test drugs have been chosen: azapropazone, dornase alpha, and losartan. MEMO has obtained facsimiles of prescriptions dispensed in Scotland between April 1994 and August 1995 for these drugs from the PPD, and allocated CHI numbers using Scottish-wide CHI databases. On the basis of up-to-date demographic details provided by the CHI, ISD

used a probability match on sex, full name, date of birth, and current and previous postcodes to identify any SMR1 hospital admission data from 1981 to June 1995 for the 14 536 patients who received these drugs. There were 43 457 hospital episodes for 10 248 of them. Analyses showed that the rate of hospital admission for upper gastrointestinal haemorrhage and perforation was 11.3 events per 1000 patient years for patients exposed to azapropazone, compared with 2.0 per 1000 patient years for unexposed patients, giving an incidence rate ratio of 5.7 (95% confidence intervals 3.5–9.3). That these results fit in with what is already known about the risk profile of azapropazone is encouraging. Analyses for the other drugs are under way. Now that this pilot study has demonstrated that patient-specific exposure datasets can be compiled for the Scottish population, the range and quality of the outcome data to which they can be linked by ISD would make drug safety studies using these resources unrivalled in detail and scope.

Studies in MEMO use highly confidential, although anonymized, medical data, and MEMO has an agreement with the Local Medical Committee of the British Medical Association never to divulge person-specific or GP-specific data, unless it is to a doctor requesting information on one of his or her own patients. All staff in MEMO sign confidentiality agreements and all databases are registered for research purposes with the Data Protection Officer. All studies use data that are anonymized by the CHI. With regard to independent ethical approval for drug safety studies, guidelines issued by the Scottish Office indicate that postmarketing surveillance studies were exempt from this requirement.¹² Research involving access to medical records without direct patient involvement is also exempt, provided that confidentiality and anonymity of patients is assured, and a senior professional person who can be disciplined by a professional body is accountable. The rationale for this is set out in a recent report into ethical issues from a Working Group of the Royal College of Physicians of London.⁴¹ However, the ethical and data protection issues are continuing to evolve and MEMO is taking steps to

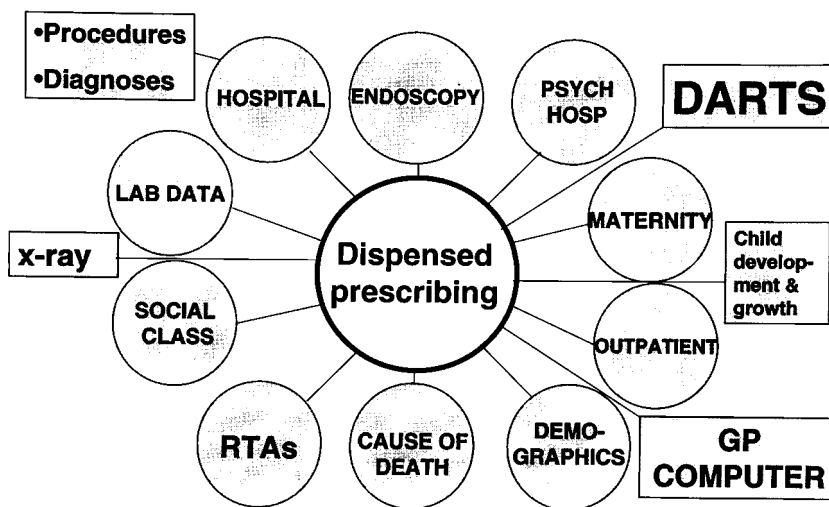


Figure 22.7. Record-linkable data that are available in MEMO.

ensure that it meets the standards in both of these areas.⁴²

In conclusion, MEMO has realized its potential as a comprehensive record linkage system that can be used for the detection and quantification of serious drug toxicity, with Figure 22.7 summarizing the record-linkable data that are now available. The next few years are likely to see increasing numbers of applications of its capabilities, along with additions to what it can accomplish.

REFERENCES

1. Beardon PHG, Brown SV, McDevitt DG. Four-year mortality among cimetidine takers in Tayside: results of a controlled study using record linkage. *Pharm Med* 1988; **3**: 333–9.
2. MacDonald TM, McDevitt DG. The Tayside Medicines Monitoring Unit (MEMO). In: Strom BL, ed., *Pharmacoepidemiology 2nd edn*. Chichester: Wiley, 1994; 245–55.
3. International Classification of Diseases. *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death 9th rev.*, Vol. 1. Geneva: World Health Organization, 1977.
4. Office of Population Censuses and Surveys. *Tabular List of the Classification of Surgical Operations and Procedures 4th rev.*, London: HMSO, 1990.
5. ICD-10. *International Statistical Classification of Diseases and Related Health Problems 10th rev.*, Geneva: World Health Organization, 1992.
6. Chisholm J. The Read Clinical Classification. *Br Med J* 1990; **300**: 1092.
7. Morris AD, Boyle DIR, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, MacDonald TM. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record-linkage to create a diabetes register. *Br Med J* 1997; **315**: 524–8.
8. Evans JMM, McMahon AD, Steinke DT, McAlpine RR, MacDonald TM. Do H_2 -receptor antagonists cause acute pancreatitis? *Pharmacoepidemiol Drug Safety* 1988; **7**: 383–8.
9. Carstairs V. Deprivation and health in Scotland. *Health Bull (Edinb)* 1990; **48**: 162–75.
10. Evans JMM, Orr C, Duncan ID, MacDonald TM. Use of hormone replacement therapy in the community. Could this be improved? *Pharmacoepidemiol Drug Safety* 1997; **6**: S81.
11. Beardon PHG, McGilchrist MM, McKendrick AD, McDevitt DG, MacDonald TM. Primary non-compliance with prescribed medication in general practice. *Br Med J* 1993; **307**: 846–8.
12. The Scottish Office, Home and Health Department. Local Research Ethical Committees. Edinburgh: HMSO, 1992: 9.
13. Pears J, Alexander V, Alexander GF, Waugh NR. Audit of the quality of hospital discharge data. *Health Bull* 1992; **50**: 356–61.
14. Kohli HS, Knill-Jones RP. How accurate are SMRI (Scottish Morbidity Record 1) Data? *Health Bull* 1992; **50**: 14–23.
15. Park RHR, McCabe P, Russell RI. Who should log the Ships? The accuracy of Scottish Hospital Morbidity Data for Wilson's Disease. *Health Bull* 1992; **50**: 24–8.

16. Evans JMM, MacDonald TM. Misclassification and selection bias in case-control studies using an automated database. *Pharmacoepidemiol Drug Safety* 1997; **6**: 313-18.
17. Beardon PHG, Brown SV, McDevitt DG. Post-marketing surveillance: a follow-up study of morbidity associated with cimetidine using record-linkage. *Pharm Med* 1988; **3**: 185-93.
18. Beardon PHG, Brown SV, McDevitt DG. Gastro-intestinal events in patients prescribed non-steroidal anti-inflammatory drugs: A controlled study using record linkage in Tayside. *QJM* 1989; **71**: 497-505.
19. Hayes JL, Evans JMM, Lipworth BP, MacDonald TM. Potentially hazardous co-prescribing of β -adrenoceptor antagonists and agonists in the community. *Br J Gen Pract* 1996; **46**: 423-5.
20. MacDonald TM, McMahon AD, Reid IC, Fenton GW, McDevitt DG. Anti-depressant drug use in primary care: A cause for concern. *Br Med J* 1996; **313**: 860-1.
21. Steinke DT, Bain DJG, Davey PG, MacDonald TM. Practice factors that influence prescribing of antibiotic and psychotropic medication in general practice in Tayside. *J Antimicrob Ther* 2000; in press.
22. Morris AD, Boyle DIR, McMahon AD, Greene SA, MacDonald TM, Newton RW. Adherence to insulin treatment, glycaemic control and ketoacidosis in insulin dependent diabetes mellitus. *Lancet* 1997; **350**: 1505-10.
23. MacDonald TM, Beardon PHG, McGilchrist MM, Duncan ID, McKendrick AD, McDevitt DG. The risks of symptomatic vaginal candidiasis following oral antibiotic therapy. *QJM* 1993; **86**: 419-24.
24. Hawkey CJ. Non-steroidal anti-inflammatory drugs and peptic ulcers. Facts and figures multiply, but do they add up? *Br Med J* 1990; **300**: 278-84.
25. McMahon AD, Evans JMM, White G, Murray FE, McGilchrist MM, McDevitt DG, MacDonald TM. A cohort study (with re-sampled comparator groups) to measure the association between new NSAID prescribing and upper gastrointestinal haemorrhage and perforation. *J Clin Epidemiol* 1997; **50**: 351-6.
26. MacDonald TM, Morant SV, Robinson GC, Shield MJ, McGilchrist MM, White G, Murray FE, McDevitt DG. The upper gastrointestinal toxicity of non-steroidal anti-inflammatory drugs is constant with continued exposure. *Br Med J* 1997; **315**: 1333-7.
27. Carson JL, Strom BL, Soper KA, West SL, Morse L. The association of non-steroidal anti-inflammatory drugs with upper gastrointestinal tract bleeding. *Arch Intern Med* 1987; **147**: 85-8.
28. Evans JMM, McMahon AD, McGilchrist MM, White G, Murray FE, McDevitt DG, MacDonald TM. Topical non-steroidal anti-inflammatory drugs and admission to hospital for upper gastrointestinal bleeding and perforation: a record-linkage case-control study. *Br Med J* 1995; **311**: 22-6.
29. Evans JMM, McGregor E, McMahon AD, McGilchrist MM, Jones MC, White G, McDevitt DG, MacDonald TM. Non-steroidal anti-inflammatory drugs and hospitalization for acute renal failure. *QJM* 1995; **88**: 551-7.
30. Evans JMM, McMahon AD, Murray FE, McDevitt DG, MacDonald TM. Non-steroidal anti-inflammatory drugs are associated with emergency hospitalisation for colitis due to inflammatory bowel disease. *Gut* 1997; **40**: 619-22.
31. Evans JMM, McGregor A, Murray FE, Vaidya K, Morris AD, MacDonald TM. No association between non-steroidal anti-inflammatory drugs and acute appendicitis: A case-control study. *Br J Surg* 1997; **84**: 372-4.
32. Barbone F, McMahon AD, Davey PG, Morris AD, Reid IC, McDevitt DG, MacDonald TM. Association of road-traffic accidents with benodizepine use. *Lancet* 1998; **352**: 1331-6.
33. McMahon AD, Evans JMM, McGilchrist MM, McDevitt DG, MacDonald TM. Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. *Pharmacoepidemiol Drug Safety* 1998; **7**: 275-80.
34. McMahon AD, MacDonald TM. Sample size for cohort studies in pharmacoepidemiology. *Pharmacoepidemiol Drug Safety* 1997; **6**: 331-5.
35. Morris AD, Boyle DIR, McMahon AD, Pearce H, Evans JMM, Newton RW, Jung RT, MacDonald TM. ACE inhibitor use is associated with hospitalization for severe hypoglycemia in patients with diabetes. *Diabetes Care* 1997; **20**: 1363-7.
36. MacDonald TM. The economic evaluation of antibiotic therapy: relevance to urinary tract infection. *J Antimicrob Chemother* 1994; **33** (suppl. A): 137-45.
37. MacDonald TM, Collins D, McGilchrist MM, Stevens J, McKendrick AD, McDevitt DG, Davey PG. The utilisation and economic evaluation of antibiotics prescribed in primary care. *J Antimicrob Ther* 1995; **35**: 191-204.
38. Davey PG, McMahon AD, Barbone F, Gillespie WG, Rizvi KA, MacDonald TM. The effect of hip replacement on prescribing of NSAIDs, ulcer healing drugs and hospitalisation. A matched cohort study. *Pharmacoepidemiol Drug Safety* 1999; **8**: 423-31.
39. McGilchrist MM, MacDonald TM. Automatic capture of community prescribing. *Pharmacoepidemiol Drug Safety* 1996; **5**: S37.
40. McGilchrist MM, McMahon AD, MacDonald TM. A study to assess record-linkage in the Scottish population. *Pharmacoepidemiol Drug Safety* 1997; **6**: A164.

41. Diamond A, Hall S, Jay M, Laurence D, Mayon-White R, Peckham C, Rawlins M, Turner-Warwick M, Wald N. Independent ethical review of studies involving personal medical records. Report of a working group to the Royal College of Physicians Committee on Ethical Issues in Medicine. *J R College Physicians* 1994; **28**: 439–42.
42. Data Protection Act 1998. Chapter 29. London: Stationery Office.

23

The UK General Practice Research Database

LUIS A. GARCÍA RODRÍGUEZ

Centro Español de Investigación Farmacoepidemiológica (CEIFE), Madrid, Spain

SUSANNE PÉREZ-GUTTHANN

Novartis Pharmaceuticals, Global Epidemiology, Barcelona, Spain

SUSAN JICK

Boston Collaborative Drug Surveillance Program, Boston University School of Medicine, Lexington, MA, USA

INTRODUCTION

Before the advent of automated databases, the identification and appropriate followup of large cohorts of users of even a single medicine was extremely expensive and required a major administrative effort to assemble and ensure complete and accurate collection of the required followup information. The initial use of computerized information which began in the late 1970s permitted new, highly efficient research. However, the available resources had characteristics which limited the quantity and types of study that could be conducted.¹ In considering how one might generate a larger, more efficient source of directly accessible medical information, it was apparent that the UK's National Health System provided a unique medical environment within which to create an optimum computerized medical data resource. In the UK, traditionally the general

practitioner (GP) acts as a gatekeeper to services within the National Health Service. This function allows a comprehensive record keeping of clinical events of individual patients. A comprehensive record of prescriptions written, outpatient diagnoses, and referral letters to hospitals resides in the individual patient records that GPs file in their offices. A few pharmacoepidemiology studies had, in fact, already been conducted based primarily on review and abstraction of information derived directly from original paper GP medical records.² Such studies were extremely expensive and the quality and quantity of data that could be derived from the search and review of clinical records was limited. Nevertheless, such studies demonstrated that GP records contain the comprehensive medical information that is required for pharmacoepidemiology studies.

In the late 1980s, VAMP Health, a commercial company, designed and marketed a GP computer

system that allowed for comprehensive recording of medical information for individual patients on office computers. This system was based on software designed in the early 1980s by Dr Alan Dean, a practicing GP in England. The system provided for computer recording of patient demographics, all prescriptions, and all clinical diagnoses, together with considerable additional information on patient medical care. From the start, the directors of VAMP, which included Dr Alan Dean, recognized that the information recorded on GP office computers could represent a highly valuable resource for clinical research in general, and for pharmacoepidemiology research in particular.

Two steps were taken to achieve the goal of constructing an automated clinical research data resource. First, VAMP signed an agreement with approximately 1000 general practices encompassing some 3000 GPs, whereby GPs would (i) agree to be trained to enter on computer critical data items in a standard manner; (ii) provide a copy of the data to VAMP, stripped of patient identifiers; and (iii) provide anonymized photocopied referral letters required to complete clinical studies. In return, GPs were to receive compensation for their cooperation.

Second, the directors of VAMP entered into a cooperative agreement with the BCDSP, a research center with many years of experience with the use of computerized data for pharmacoepidemiology,^{1,3,4} whereby the BCDSP received the anonymized raw data generated by the GPs in order to evaluate its potential use for pharmacoepidemiology research and to conduct such research if the data quality and completeness were satisfactory.

The initial task to be accomplished was to merge the information from hundreds of participating general practitioners into a single file. In order to accomplish this crucial first step, it required that the participating GPs enter the required data in a standard manner. Initially there were approximately 600 general practices, caring for about four million patients, who fulfilled this requirement; a single file encompassing these practices was created. The next step was to reorganize the files into a form that was efficient and designed for

pharmacoepidemiology studies. Efficient access also required that the codes used by VAMP to identify medicines and diagnoses, which numbered in the tens of thousands, be mapped onto more efficient coding schemes. For instance, at the BCDSP, the original drug dictionary was mapped to a unique BCDSP drug coding system and the diagnosis dictionary was mapped onto the International Classification of Diseases (ICD) coding system. The original coding systems can be used to obtain particular details of individual products for the drug codes such as strength and route of administration, or to obtain more specific diagnoses.

The next and critical step was to evaluate the quality and completeness of the recorded data. In order to determine the quality of the diagnosis information, the BCDSP conducted extensive studies involving thousands of patients, in which the diagnoses present on photocopies of referral letters from hospital consultants were compared to the diagnoses recorded on computer (see below).^{5,6}

Since 1994, the database has been known as the General Practice Research Database (GPRD), has belonged to the UK Department of Health, and has been maintained by the Office of National Statistics (ONS).⁷ The GPRD contains computerized information entered by GPs after a trial period of data entry and quality training. For a limited number of practices, prospective data collection is available from as early as 1987. Starting in 1991, most practices participating in the GPRD have been providing data with the quality and completeness required for pharmacoepidemiology research projects. Currently, around 1500 general practitioners, with a population coverage in excess of 3 million, systematically provide their computerized medical data anonymously to the ONS. Upon receiving data from the GPs, the ONS personnel organize this information and perform a series of quality checks. Access to these data has also been broadened to include others, beyond just the BCDSP. Data for research projects can be obtained from the ONS after review of research protocols by a Scientific and Ethical Advisory Board (SEAG). The main task of SEAG is to ensure that scientific and ethical standards are maintained, including anon-

ymity. All information remains strictly anonymized, as no patient, general practitioner, or practice identifiers are present in the data sent to investigators. The costs involved in the collection of the data and the present scheme of funding of the GPRD have been reviewed elsewhere.⁸

DESCRIPTION

The GPRD data are recorded using four data screens or files. These include the registration file, the drug file, the events (diagnoses, procedures) file, and the “findings” file.

REGISTRATION FILE

This file contains information regarding the registration status of all individuals covered in the GPRD data files. As with other GPRD files, this one contains a unique patient identification number that enables information from all files to be linked. There is also a number that identifies the practice (though the practice identity is confidential and known only to people in the MCA who maintain the data, and to people working in Alan Dean’s EPIC Research Unit who provide medical record retrieval services). Finally, there is a family identification number that enables members of the same family to be linked (e.g., mothers and babies etc.). Other information includes year of birth (and month for those under age 5), gender, date of registration with the current general practice, most recent registration status (permanent, transferred out, died), and date of death or departure from the practice (where applicable). All computerized data are processed to ensure complete patient and GP confidentiality and anonymity, when data are transmitted to researchers.

DRUG FILE

This file contains detailed information on all drugs prescribed by the GP. Drugs are coded with the prescription pricing authority (PPA) coding system, with a specific code for each commercial preparation. Details include the date of each drug prescription, the precise drug formulation and

strength, the quantity of drug prescribed, and the dosing instructions. In addition, the indication for treatment is required for all new courses of therapy. This is done by cross-referencing prescriptions against medical events on the same date. GPs participating in the GPRD data scheme are required to record all vaccinations administered as well as contraception prescribed. In certain instances, a specialist may initiate a course of treatment; therefore, this first prescription is not recorded in the database. Refills for subsequent prescriptions are then written by GPs and recorded in the database. It is worth noting that when patients are seen by consultants or in hospital, future drug therapy is directed through the GP and is thus captured by the database.

EVENT FILE

This file contains all clinically relevant patient diagnoses, both inpatient and outpatient, along with the date of the event. In addition to usual care in the GP office, GPs are required to enter the indication for any new drug therapy, as noted above, as well as all diagnoses resulting from hospitalizations, consultations, or emergency medical care. There is also a field that indicates, for each diagnostic code, where the care was received (in the GP’s office, the hospital, an outpatient consultant, emergency care, etc.). Diagnoses or events are coded using the OXMIS (Oxford medical indexing system) coding system.⁹ This dictionary was originally created in the 1960s to specifically address the coding needs of physicians in a general practice setting. The coding is loosely based on the ICD coding but it has seven digits including alpha and numeric characters to provide more detail in the coding. This data file also contains a comment field that allows the GP to enter up to 19 characters of free text for each diagnosis entry. These comments often provide useful additional details about the diagnosis coded.

Because the GP is the primary care giver for all patients in the National Health Service, all consultants are required to send a letter to the GP whenever a patient is seen in hospital or by an outpatient specialist, describing the relevant

clinical events and final diagnoses. The contents of these letters are then entered into the computer file by the GP. The details of a patient's hospital or specialist clinical history are not recorded on the computer, although the final diagnoses are.

FINDINGS FILE

This file contains patient information in areas such as blood pressure, height, weight, smoking status, alcohol use, contraceptive use, cervical smears, immunizations, medical procedures, laboratory tests, results, and social factors. While the GPs are not required to update all of these data fields, information on characteristics such as height, weight, and smoking have been used repeatedly and are available on greater than 70% of the population.

MEDICAL RECORDS

While most information is available on the computerized medical files, it is essential to have access to the original medical records in order to validate recorded diagnoses and to obtain additional detailed information about hospitalized and referred medical events. It is essential, in most studies, to know whether a diagnosis is a first time (incident) diagnosis and the precise date of onset, whether or not there was some exclusion criteria not recorded on the database, and whether the diagnosis was confirmed by certain tests or procedures. These details cannot always be ascertained with complete confidence from the computer record. As noted above, access to consultant medical records is provided by Alan Dean's EPIC Research Unit.

STRENGTHS

POPULATION-BASED DATA

As the UK National Health System provides universal coverage, no segment of the population is excluded from the GPRD. The nature of the data makes it possible to link mother and offspring

information to study human reproductive epidemiology, and in particular the teratogenicity of drugs used during pregnancy.¹⁰ The geographical distribution of the practices participating in the GPRD is representative of the UK population, apart from small variations among regions. Recent comparisons of age and sex distributions with the National Population Census have shown these to be very similar.¹¹

POPULATION LARGE ENOUGH TO STUDY RARE DISEASES

Over 3 million persons are included in the GPRD at any one time. For most practices, data have been entered prospectively on computer for a period exceeding seven years. In addition, many GPs have entered at their discretion historical medical data preceding prospective data recording. Data collected to date represent some 25 million person years of information. This has allowed researchers to study rare outcomes, with an incidence rate of less than one per 10 000 persons per year, as well as chronic disorders and conditions with short or medium term incubation periods.

QUALITY OF THE DIAGNOSIS INFORMATION

One of the primary purposes of GPs' data recording is the daily clinical management of their patients; there is no separate medical record. As such, to the degree the care was provided by the GP, it should be recorded very accurately. The only potential exception would be the consultant and hospital records, which need to be entered into the data manually. However, validation studies of the GPRD have documented the recording of medical data in the general practitioners' computers to be near complete.^{5,6,12} In several independent studies, hospital consultant letters in the manual records of GPs were photocopied and the consultant's clinical diagnoses were compared to those recorded on the GPs computer. Agreement between the consultants' letters and the GP computer records was over 87% (1038/1191) in a

1991 study, over 90% (121/126) in a 1992 study, and over 95% (539/553) in a 1994 study.^{5,6,12} In addition, numerous research papers requiring validation of coded diagnostic outcomes have shown good agreement between the computer recorded diagnosis and the diagnosis on the written clinical records (see appendix).

Regarding outpatient drug information, the prescriptions are issued to the patients directly from the computer, insuring that virtually all outpatient prescriptions are captured by the database and that the prescription information is accurate. GPRD also therefore includes the dosage instruction written by GPs. There is over 90% concordance between the prescriptions recorded on the computer and those recorded at the UK Prescription Pricing Authority.^{13,14}

ACCESS TO ORIGINAL MEDICAL RECORDS

One of the *sine qua non* conditions to carry out valid research using automated databases is the possibility of obtaining copies of original medical records and death certificates. Access to original records is needed in order to confirm the diagnosis initially identified through a computer search, as well as to abstract additional information on past history and laboratory findings. To date, studies performed with the GPRD in which medical records were requested have resulted in response rates well over 80%, and in many 90% or above. Due to the excellent collaboration of the participating GPs, around 80% of medical records are received within three months of the initial request date. Another unique additional feature of the GPRD is the possibility of sending the GP project-specific questionnaires requesting information that is usually not recorded in the computer files. In some rare instances, questionnaires have also been sent to patients through their GPs, once consent was obtained from both GPs and patients. As previously mentioned, all computer and paper based information is anonymized before being sent to researchers.

WEAKNESSES

COMPLETENESS OF DATA

As noted above, the GPRD file represents the primary medical record for the GP, and as such the GP data are likely to be very complete. The same is not true of diagnoses provided by specialists and care provided in hospitals. One of the requirements to become a participant in the GPRD is to agree to record all referrals to specialists as well as the information resulting from these visits. Three validation studies have documented the completeness of this information.^{5,6,12} In these studies, researchers reviewed information from anonymized photocopied consultant referral letters present in the offices of GPs and compared them to the diagnoses recorded in the computer files. The concordance for clinically relevant outcomes was between 90 and 95%. While this clearly is reassuring, 5–10% missing data could be critically important when studying outcomes that occur uncommonly. Further, these studies do not evaluate the completeness of hospital outcomes, nor quantify the degree to which the GPs do not have in their files consultant referral letters corresponding to care provided to their patients.

ADDITIONAL COMPUTERIZED INFORMATION

At least in the area of pharmacoepidemiology, the essential data for studies are demographic information, outpatient prescriptions, and hospital based diagnoses. However, in this and other research areas, a number of other variables would be welcome, such as smoking habits, weight, height, life style (diet, exercise) and socioeconomic and marital status. Some of these are already being recorded in the GPRD, but not yet routinely. Currently, data on smoking, weight, and height are available for over 70% of the population, but this leaves nearly 30% of the data missing. Socioeconomic status information is not recorded for each patient, although this information is available for the practice and that could potentially be used as a surrogate. A limitation in reproductive

epidemiology is the relatively low recording of last menstrual period in the pregnant woman, and weight and length at birth in the offspring. Also, impacting specific areas of research are low recording levels of data such as date of menarche and menopause, family medical history, and historical information on surgical interventions and chronic conditions. Other useful information would be data on over-the-counter medication (by asking the patient at each visit about over-the-counter drug use), in-hospital drug use, and other health care treatments delivered outside the GP practice (e.g., chemotherapy, PUVA). Finally, direct economic information on health care resource utilization is not present in the GPRD.

LINKAGE TO OTHER HEALTH CARE DATABASES

Linkage to other existing health care automated databases would be welcome to facilitate followup and diagnostic validation processes. One of the most desirable would be a linkage to hospital databases containing information on discharge diagnoses and procedures performed. Other useful linkages would be to cancer registries, congenital malformation reporting systems, and laboratory test result databases. In principle, these links could be arranged.

ADDITIONAL RESEARCH UNITS

The complexities of observational data and in particular GPRD data require an extended period of learning before embarking on studies using the GPRD as the primary source of information. Experience with the GPRD is still limited and restricted to a handful of research units around the world. Also, input from these researchers could help to improve certain aspects of future data collection.

PARTICULAR APPLICATIONS

The appendix lists a compilation of 80 papers that have been published to date using the GPRD. The majority are pharmacoepidemiology studies, cov-

ering therapeutic areas such as nonsteroidal anti-inflammatory drugs, hormone replacement therapy, anti-infective agents, oral contraceptives, antihypertensive drugs, acid suppressing drugs, antidepressants, anticonvulsants, and drugs to treat diseases such as asthma and diabetes. The clinical outcomes of these studies included liver disorders, upper gastrointestinal bleeding and perforation, metabolic and endocrine disorders, seizures, suicide, blood dyscrasias, severe cutaneous reactions, ocular disorders, myocardial infarction, neoplasia, sudden unexplained death, congenital malformations, venous thromboembolism, renal disorders, and bacterial infections. The database has been less frequently used for studies of drug utilization, natural history of disease, pharmacoconomics, and health care utilization.¹⁵ However, a few studies are available in these areas: resources used by schizophrenic patients, evaluation of prevention strategies in the general practice setting, resources needed for the care of patients with eating disorders, use of influenza vaccines, and natural history and drug utilization in asthma and diabetes mellitus. As examples, we will briefly review three of the pharmacoepidemiology studies performed with the GPRD: one that quantified the risk of acute hepatic injury in patients using the combination of amoxicillin with clavulanic acid, another one that compared the risk of developing cancer in users of different antihypertensive agents, and lastly a study that evaluated the risk of heart valve disorders in users of diet medicines.

LIVER INJURY AND ANTIBIOTICS

Several studies have been performed with the GPRD on the epidemiology of drug induced liver injury.^{16,17} Among antibiotics incriminated as hepatotoxic, use of the combination of amoxicillin and clavulanic acid has been the source of numerous case reports since the late 1980s. A large cohort study including over 90 000 users of the combination of amoxicillin and clavulanic acid and 360 000 users of amoxicillin alone was identified with the GPRD, and followed for a two year period.¹⁶ The objective of the study was to quantify and compare the risk of acute liver injury among users of these two antibiotics. After

review of the original medical records, 35 cases met all the case definition criteria. The absolute risks of acute liver injury associated with the combination of amoxicillin and clavulanic acid, and amoxicillin alone, were 1.7 and 0.3 per 10 000 prescriptions, respectively. The data also made it possible to characterize the groups at highest risk, namely elderly patients receiving more than one course of amoxicillin and clavulanic acid combination therapy, with a risk of developing acute liver injury of greater than ten per 10 000. The clinical features of liver injury could be evaluated in detail based on the information from the medical records. This showed that the liver injury was predominantly cholestatic, and occasionally presented after the end of the combination therapy, confirming earlier individual case reports. Data from this study formed part of the evidence for revised conditions of use of the combination of amoxicillin and clavulanic acid, in the UK and abroad.¹⁸

CALCIUM CHANNEL BLOCKERS AND CANCER

An association was recently reported between the use of calcium channel blockers (CCBs) and an increased risk of all cancer. Using the GPRD, a study was conducted to assess whether cancer was associated with CCB use.¹⁹ The source cohort was restricted to hypertensive patients using β -blockers only, inhibitors of angiotensin converting enzyme (ACE) only, or CCBs only. A nested case-control analysis was performed within this cohort. A total of 446 cases of cancer and 1750 controls were analyzed. After adjustment for smoking, body mass index, change of antihypertensive medication, duration of hypertension, and diuretic use, the odds ratio estimates for all cancers combined were 1.3 (95% CI 1.0–1.6) for CCBs and 0.8 (95% CI 0.6–1.2) for ACE inhibitors, relative to β -blockers. The corresponding odds ratio among patients older than 70 years was 1.2. No duration effect was found among users of CCBs, and a small dose effect was observed. Subgroup analyses by specific cancers did not show an increased risk for any particular location associated with use of CCBs. The authors concluded that no evidence of a material increase in the risk of any cancer or of

any particular cancers was found associated with use of calcium channel blockers, relative to use of β -blockers. Subsequently, two other studies also failed to demonstrate an excess risk among users of CCBs.^{20,21} However, additional research is still needed to resolve the controversy.

DIET MEDICINES AND HEART VALVE DISORDERS

Case reports have suggested that a combination of the appetite suppressants fenfluramine and phentermine is associated with an increased risk of cardiac valve regurgitation. There were also reports of valvular disorders in persons taking fenfluramine or dexfenfluramine alone. Using the GPRD, a population based followup study and a nested case-control analysis were conducted on 6532 subjects who received dexfenfluramine, 2371 who received fenfluramine, and 862 who received phentermine to assess the risk of a subsequent clinical diagnosis of a valvular disorder of uncertain origin.²² For comparison, a group of 9281 obese subjects who had not taken appetite suppressants who were matched to the treated subjects for age, sex, and weight were identified. All subjects were free of diagnosed cardiovascular disease at the start of followup. The average duration of followup for all subjects was about four years.

There were 11 cases of newly diagnosed idiopathic valvular disorders, five after the use of dexfenfluramine and six after the use of fenfluramine. There were six cases of aortic regurgitation, two cases of mitral regurgitation, and three cases of combined aortic and mitral regurgitation. There were no cases of idiopathic cardiac valve abnormalities among the subjects who had not taken appetite suppressants or among those who took only phentermine. The five year cumulative incidence of idiopathic cardiac valve disorders was 0 per 10 000 subjects (95% CI 0–15.4) among those who had not taken appetite suppressants; 0 per 10 000 subjects (95% CI 0–76.6) among those who took phentermine alone; 7.1 per 10 000 subjects (95% CI 3.6–17.8) among those who took either fenfluramine or dexfenfluramine for less than four months ($p = 0.02$ versus those who

had not taken appetite suppressants); and 35.0 per 10 000 subjects (95% CI 16.4–76.2) among those who received either of these medications for four or more months ($p < 0.001$ versus those who had not taken appetite suppressants). The authors concluded that the use of fenfluramine or dexfenfluramine, particularly for four months or longer, is associated with an increased risk of newly diagnosed cardiac valve disorders, particularly aortic regurgitation.

THE FUTURE

In the recent history of pharmacoepidemiology research in Europe, the UK General Practice Research Database is the single largest source of published studies. GPRD data have been used extensively in the area of pharmacoepidemiology, contributing on several occasions to decision making by regulatory authorities and industry. However, there are still some areas where the GPRD has been underused, such as drug utilization, natural history of disease, pharmacoecconomics, and health care utilization research.

Numerous studies have documented the completeness and high quality of the information recorded in the GPRD, as well as the excellent access to original medical records and collaboration of general practitioners. Welcome improvements would be the recording of additional information on life style habits and linkage to other medical databases.

The General Practice Research Database has proved that valuable data can be collected in a general practice setting. The potential of this rich computerized database has not yet been fully exploited. Hopefully, the coming years will witness new projects and expanding applications of this database. This experience should also serve as a stimulus to efforts to generate similar population based data in other countries.

ACKNOWLEDGEMENTS

The authors thank Dr Hershel Jick for his valuable comments on the manuscript.

APPENDIX: PUBLICATIONS BASED ON THE GPRD

1. Jick H, Hall GC, Dean AD, Jick SS, Derby LE. A comparison of the risk of hypoglycemia between users of human and animal insulins 1. Experience in the United Kingdom. *Pharmacotherapy* 1990; **10**: 395–7.
2. Teeling-Smith G. Low-level prescribing not necessarily good prescribing. *Prescriber* 1990; July: 37–42.
3. Lawrence M, Coulter A, Jones L. A total audit of preventive procedures in 45 practices caring for 430, 000 patients. *Br Med J* 1990; **300**: 1501–3.
4. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *Br Med J* 1991; **302**: 766–8.
5. Tallis R, Hall G, Craig I, Dean A. How common are epileptic seizures in old age? *Age Ageing* 1991; **20**: 442–8.
6. Jick SS, Jick H, Knauss TA, Dean AD. Antidepressants and convulsions. *J Clin Psychopharmacol* 1992; **12**: 241–5.
7. Derby LE, Jick H, Dean AD. Antidepressant drugs and fatal suicide. *J Clin Psychopharmacol* 1992; **12**: 235–40.
8. Jick H, Derby LE, García Rodríguez LA, Jick SS, Dean AD. Liver disease associated with diclofenac, naproxen and piroxicam. *Pharmacotherapy* 1992; **12**: 207–12.
9. Orlander JD, Jick SS, Dean AD, Jick H. Urinary tract infections and estrogen use in older women. *J Am Geriatr Soc* 1992; **40**: 817–20.
10. Jick H, Ulcickas M, Dean A. Comparison of frequencies of suicidal tendencies among patients receiving fluoxetine, lofepramine, mianserin or trazodone. *Pharmacotherapy* 1992; **12**: 451–4.
11. Mannino S, García Rodríguez LA, Jick SS. NSAIDs, quinolones and convulsions: an epidemiologic approach. *Post Marketing Surveillance* 1992; **6**: 119–28.
12. Jick H, Terris BZ, Derby LE, Jick SS. Further validation of information recorded on a general practitioner based computerized data resource in the United Kingdom. *Pharmacoepidemiol Drug Safety* 1992; **1**: 347–9.
13. Wild RN. Micturin and torsades de pointe—experience of a post—marketing alert. *Pharmacoepidemiol Drug Safety* 1992; **1**: 147–50.
14. Derby LE, Jick H, Henry DA, Dean AD. Cholestatic hepatitis associated with flucloxacillin. *Med J Aust* 1993; **158**: 596–600.
15. Derby LE, Jick H, Henry DA, Dean AD. Erythromycin-associated cholestatic hepatitis. *Med J Aust* 1993; **158**: 600–2.

16. Jick H, Derby LE, García Rodríguez LA, Jick SS, Dean AD. Nonsteroidal anti-inflammatory drugs and certain rare serious adverse effects. A cohort study. *Pharmacotherapy* 1993; **13**: 212–17.
17. Derby LE, Pérez Gutthann S, Jick H, Dean AD. Liver disorders in patients receiving chlorpromazine or isoniazid. *Pharmacotherapy* 1993; **13**: 353–8.
18. Jick SS, Jick H, Dean AD. A follow-up safety study of ciprofloxacin users. *Pharmacotherapy* 1993; **13**: 461–4.
19. Sleator DJD. Towards accurate prescribing analysis in general practice: accounting for the effects of practice demography. *Br J Gen Pract* 1993; **43**: 102–6.
20. Nazareth I, King M, Haines A, See Tai S, Hall G. Care of schizophrenia in general practice. *Br Med J* 1993; **307**: 910.
21. Nazareth I, King M, Haines A, Rangel L, Myers S. Accuracy of diagnosis on general practice computer system. *Br Med J* 1993; **307**: 32–4.
22. Lis Y, Spitzer WO, Mann RD, Cockburn I, Chukwujindu J, Thorogood M, Evans S, Suissa S, Leloir J. A concurrent cohort study of oral contraceptive users from the VAMP research bank. *Pharmacoepidemiol Drug Safety* 1993; **2**: 51–3.
23. García Rodríguez LA, Williams R, Derby LE, Dean AD, Jick H. Acute liver injury associated with nonsteroidal anti-inflammatory drugs and the role of risk factors. *Arch Intern Med* 1994; **154**: 311–16.
24. García Rodríguez LA, Jick H. Comparison of the risk of gynaecomastia associated with cimetidine, omeprazole and other antiulcer medications. *Br Med J* 1994; **308**: 503–6.
25. García Rodríguez LA, Jick H. The risk of upper gastrointestinal bleeding and/or perforation associated with individual nonsteroidal anti-inflammatory drugs. *Lancet* 1994; **343**: 769–72.
26. Jick H, Derby LE, Dean AD, Henry DA. Flucloxacillin and cholestatic hepatitis. *Med J Aust* 1994; **160**: 525.
27. Mann RD. The VAMP system and the detection of risk. Some examples. *Pharmacoepidemiol Drug Safety* 1994; **3**: 313–19.
28. Mann RD, Lis Y, Chukwujindu J, Chanter DO. A study of the association between hormone replacement therapy, smoking and the occurrence of myocardial infarction in women. *J Clin Epidemiol* 1994; **47**: 307–12.
29. Jick SS, Dean AD, Jick H. Antidepressants and suicide. *Br Med J* 1995; **310**: 215–18.
30. Jick H, Withers JM, Dean AD. Haemophilus influenza vaccine. *Br J Gen Pract* 1995; **45**: 107.
31. Jick H, Myers MW, Dean AD. Sulfasalazine and mesalazine associated blood disorders. *Pharmacotherapy* 1995; **15**: 176–81.
32. Jick H, Derby, LE. Is co-trimoxazole safe? *Lancet* 1995; **345**: 1118–19.
33. Jick H, Derby, LE. A large population-based follow-up study of trimethoprim/sulfamethoxazole, trimethoprim, and cephalexin users for certain uncommon serious drug toxicity. *Pharmacotherapy* 1995; **15**: 428–32.
34. Jick H. *A Major Resource for Drug Safety Studies. The General Practice Research Database (Formerly VAMP Research)*. Carshalton: Centre for Medicines Research, 1995.
35. Jick H, Jick SS, Gurewich V, Myers MW, Vasilakis C. Risk of idiopathic cardiovascular death and nonfatal venous thromboembolism in women using oral contraceptive with differing progestagen components. *Lancet* 1995; **346**: 1589–93.
36. García Rodríguez LA, Mannino S, Wallander MA. Ocular safety of antiulcer drugs. *Lancet* 1995; **345**: 1059–60.
37. Lancaster T, Surman G, Lawrence M, Mant D, Vessey M, Thorogood M, Yudkin P, Daly E. Hormone replacement therapy: characteristics of users and non-users in a British general practice cohort identified through computerised prescribing records. *J Epidemiol Community Health* 1995; **49**: 389–94.
38. Lloyd DCEF, Harris CM, Roberts DJ. Specific therapeutic group age–sex related prescribing units (STAR-PUs): weightings for analysing general practices' prescribing in England. *Br Med J* 1995; **311**: 991–4.
39. Jick H, Jick SS, Myers MW, Vasilakis C. Risk of acute myocardial infarction and low-dose combined oral contraceptives. *Lancet* 1996; **347**: 627–8.
40. Derby LE, Tennis P, Jick H. Sudden unexplained death among subjects with refractory epilepsy. *Epilepsia* 1996; **37**: 931–5.
41. Jick H, Derby LE, Gurewich V, Vasilakis C. The risk of myocardial infarction associated with anti-hypertensive drug treatment in persons with uncomplicated essential hypertension. *Pharmacotherapy* 1996; **16**: 321–6.
42. García Rodríguez LA, Stricker BH, Zimmerman HJ. Risk of acute liver injury associated with the combination of amoxicillin and clavulanic acid. *Arch Intern Med* 1996; **156**: 1327–32.
43. Turnbull SJ, Ward A, Treasure JL, Jick H, Derby LE. The demand for eating disorder care. An epidemiological study using the General Practice Research Database. *Br J Psychiatry* 1996; **169**: 705–12.
44. García Rodríguez LA, Mannino S, Wallander MA, Lindblom B. A cohort study on the ocular safety of anti-ulcer drugs. *Br J Clin Pharmacol* 1996; **42**: 213–16.
45. Henry D, Lim L, García Rodríguez LA, Pérez Gutthann S, Carson JL, Griffin M, Savage R, Logan R, Moride Y, Hawkey C, Hill S, Fries JT. Variability in risk of gastrointestinal complications

- with individual non-steroidal anti-inflammatory drugs: results of collaborative meta-analysis. *Br Med J* 1996; **312**: 1563–6.
46. Donovan S. The use of the General Practice Research Database (GPRD) to examine potential links between antidepressant medication and the incidence of suicide. *Drug Inf J* 1996; **30**: 785–92.
 47. García Rodríguez LA, Wallander MA, Stricker BH. The risk of acute liver injury associated with cimetidine and other acid-suppressing anti-ulcer drugs. *Br J Clin Pharmacol* 1997; **43**: 183–8.
 48. Jick S. Ciprofloxacin safety in a pediatric population. *Pediatr Infect Dis J* 1997; **16**: 130–4.
 49. Jick SS, Terris BZ. Anticonvulsants and congenital malformations. *Pharmacotherapy* 1997; **17**: 561–4.
 50. Jick H, Vasilakis C, Derby LE. Antihypertensive drugs and fatal myocardial infarction in persons with uncomplicated hypertension. *Epidemiology* 1997; **8**: 446–8.
 51. Pérez Gutthann S, García Rodríguez LA, Castellsague J, Duque Oliart A. Hormone replacement therapy and risk of venous thromboembolism: population based case-control study. *Br Med J* 1997; **314**: 796–800.
 52. Jick H, Jick SS, Derby LE, Vasilakis C, Myers MW, Meier CR. Calcium channel blockers and risk of cancer. *Lancet* 1997; **349**: 525–8.
 53. Meier CR, Jick H. Drug use and pulmonary death rates in increasingly symptomatic asthmatics in the U.K. *Thorax* 1997; **52**: 612–17.
 54. García Rodríguez LA, Ruigómez A. Gastric acid, acid-suppressing drugs and bacterial gastroenteritis: how much of a risk? *Epidemiology* 1997; **8**: 571–4.
 55. García Rodríguez LA, Ruigómez A, Jick H. A review of epidemiologic research on drug-induced acute liver injury using the General Practice Research Database in the U.K. *Pharmacotherapy* 1997; **17**: 721–8.
 56. García Rodríguez LA, Wallander MA, Johansson S, Björk S. Renal disease and acid-suppressing drugs. *Pharmacoepidemiology and Drug Safety* 1997; **6**: 247–51.
 57. Jick H. A database worth saving. *Lancet* 1997; **350**: 1045–6.
 58. Meier CR, Jick H. Omeprazole, H₂-blockers and polyarthralgia. *Br Med J* 1997; **315**: 1283.
 59. Meier CR, Jick H. Omeprazole, other antiulcer drugs and newly diagnosed gout. *Br J Clin Pharmacol* 1997; **44**: 175–8.
 60. Walker AM, Szneke P, Bianchi LA, Field LG, Sutherland LR, Dreyer NA. 5-aminosalicylates, sulfasalazine, steroid use and complications in patients with ulcerative colitis. *Am J Gastroenterol* 1997; **92**: 816–20.
 61. Lancaster T, Jick H, Swart AM. Risk of serious haematological toxicity associated with the use of chloramphenicol eye drops. *Br Med J* 1998; **316**: 667.
 62. Derby LE, Jick H. Prior appendectomy protects against ulcerative colitis. *Epidemiology* 1998; **9**: 205–7.
 63. Meier CR, Sturkenboom MCJM, Cohen AS, Jick H. Postmenopausal estrogen replacement therapy and the risk of developing systemic lupus erythematosus or discoid lupus. *J Rheumatol* 1998; **25**: 1515–9.
 64. Jick SS, Vasilakis C, Martinez C, Jick H. A study of the relation of exposure to quinolones and suicidal behavior. *Br J Clin Pharmacol* 1998; **45**: 77–81.
 65. García Rodríguez LA, Ruigómez A, Hasselgren G, Wallander MA, Johansson S. Comparison of mortality from peptic ulcer bleed between patients with or without peptic ulcer antecedents. *Epidemiology* 1998; **9**: 452–6.
 66. Meier CR, Jick H. Tamoxifen and risk of idiopathic venous thromboembolism. *Br J Clin Pharmacol* 1998; **45**: 608–12.
 67. Jick SS, Vasilakis C, Jick H. Changes in subsequent usage, pregnancies, and pregnancy terminations among oral contraceptive users following the October 1995 CSM warning on oral contraceptives containing third generation progestagens. *Lancet* 1998; **351**: 1404–5.
 68. Meier CR, Jick SS, Derby LE, Vasilakis C, Jick H. Acute respiratory tract infections and risk of first-time acute myocardial infarction. *Lancet* 1998; **351**: 1467–71.
 69. Myers MW, Jick H. Terfenadine and risk of acute liver disease. *Br J Clin Pharmacol* 1998; **46**: 251–3.
 70. Jick H, Derby LE, Vasilakis C, Fife D. The risk of seizures associated with tramadol. *Pharmacotherapy* 1998; **18**: 607–11.
 71. Jick H, Vasilakis C, Weinrauch LA, Meier CR, Jick SS, Derby LE. A population-based study of users of dexfenfluramine, fenfluramine and phentermine to estimate the risk of clinically diagnosed idiopathic valvular heart disorders. *New Engl J Med* 1998; **339**: 719–24.
 72. Sturkenboom MC, Meier CR, Jick H, Stricker BH. Minocycline and lupuslike syndrome in acne patients. *Arch Intern Med* 1999; **159**: 493–7.
 73. Ruigómez A, García Rodríguez LA. Presence of diabetes related complication at the time of NIDDM diagnosis: an important prognostic factor. *Eur J Epidemiol* 1998; **14**: 439–45.
 74. Blackburn SC, Duque Oliart A, García Rodríguez LA, Pérez Gutthann S. Antiepileptics and blood dyscrasias: a cohort study. *Pharmacotherapy* 1998; **18**: 1277–83.
 75. Meier CR, Derby LE, Jick SS, Vasilakis C, Jick H. Antibiotics and risk of subsequent first-time acute myocardial infarction. *J Am Med Assoc* 1999; **281**: 427–31.
 76. Jick SS. Pregnancy outcomes after maternal exposure to fluconazole exposure in pregnancy. *Pharmacotherapy* 1999; **19**: 221–2.

77. Derby LE, Myers MW, Jick H. Use of dexfenfluramine, fenfluramine and phentermine and the risk of stroke. *Br J Clin Pharmacol* 1999; **47**: 565–9.
78. García Rodríguez LA, Ruigómez A. Increased risk of irritable bowel syndrome after bacterial gastroenteritis: cohort study. *Br Med J* 1999; **318**: 565–6.
79. García Rodríguez LA, Ruigómez A. Secondary prevention of upper gastrointestinal bleeding associated with maintenance acid-suppressing treatment in patients with peptic ulcer bleed. *Epidemiology* 1999; **10**: 228–32.
80. de Abajo FJ, García Rodríguez LA. Risk of ventricular arrhythmias associated with nonsedating antihistamine drugs. *Br J Clin Pharmacol* 1999; **47**: 307–13

REFERENCES

1. Jick H. *A Major Resource for Drug Safety Studies. The General Practice Research Database (Formerly VAMP Research)*. Carshalton: Centre for Medicines Research, 1995.
2. Colin-Jones DG, Langman MJS, Lawson DH, Vessey MP. Postmarketing surveillance of the safety of cimetidine; mortality during second, third and fourth years of follow-up. *Br Med J* 1985; **291**: 1084–8.
3. Cohen MR, Weaver J. A compilation of abstracts and an index of articles published by the Boston Collaborative Drug Surveillance Program 1966–1991. *Hosp Pharm* 1992; **27** S3–S55.
4. Jick H, Vessey M. Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiol* 1978; **107**: 1–7.
5. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *Br Med J* 1991; **302**: 766–8.
6. Jick H, Terris BZ, Derby LE, Jick SS. Further validation of information recorded on a general practitioner based computerized data resource in the United Kingdom. *Pharmacoepidemiol Drug Safety* 1992; **1**: 347–9.
7. Hollowell J. The General Practice Research Database: quality of morbidity data. *Population Trends* 1997; Spring: 36–40.
8. Walley T, Mantgani A. The UK General Practice Research Database. *Lancet* 1997; **350**: 1097–9.
9. Perry J (ed). *OXMIS problem codes for primary medical care*. Oxford: Headington, 1978.
10. Jick SS, Terris BZ. Anticonvulsants and congenital malformations. *Pharmacotherapy* 1997; **17**: 561–4.
11. Key Health Statistics from General Practice. *Studies on Medical and Population Subjects*, No. 60. London: Stationery Office, 1996.
12. Van Staa TP, Abenheim L. The quality of information recorded on a UK database of primary care records: a study of hospitalizations due to hypoglycemia and other conditions. *Pharmacoepidemiol Drug Safety* 1994; **3**: 15–21.
13. Hollowell J. *GPRD—Scope and Quality of Data*. London: OPCS, 1994.
14. Nazareth I, King M, Haines A, Rangel L, Myers S. Accuracy of diagnosis on general practice computer system. *Br Med J* 1993; **307**: 32–4.
15. García Rodríguez LA, Pérez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol* 1998; **45**: 419–25.
16. García Rodríguez LA, Stricker BH, Zimmerman HJ. Risk of acute liver injury associated with the combination of amoxicillin and clavulanic acid. *Arch Intern Med* 1996; **156**: 1327–32.
17. García Rodríguez LA, Ruigómez A, Jick H. A review of epidemiologic research on drug-induced acute liver injury using the General Practice Research Database in the U.K. *Pharmacotherapy* 1997; **17**: 721–8.
18. MCA/CSM. Revised indications for co-amoxiclav (Augmentin). *Current Problems Pharmacovigilance* 1997; **24**: 4.
19. Jick H, Jick SS, Derby LE, Vasilakis C, Myers MW, Meier CR. Calcium channel blockers and risk of cancer. *Lancet* 1997; **349**: 525–8.
20. Olsen JH, Sorensen HT, Frits S, McLaughlin JK, Stefensen FH, Nielsen GL, Andersen M, Fraumeni JF, Olsen JJ. Cancer risk in users of calcium channel blockers. *Hypertension* 1997; **29**: 1091–4.
21. Rosenberg L, Rao RS, Palmer JR, Strom BL, Stolley PD, Zaubler AG, Warshauer E and Shapiro S. Calcium channel blockers and the risk of cancer. *J Am Med Assoc* 1998; **279**: 1000–4.
22. Jick H, Vasilakis C, Weinrauch LA, Meier CR, Jick SS, Derby LE. A population-based study of users of dexfenfluramine, fenfluramine and phentermine to estimate the risk of clinically diagnosed idiopathic valvular heart disorders. *New Engl J Med* 1998; **339**: 719–24.

24

Other Approaches to Pharmacoepidemiology Studies

BRIAN L. STROM

Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine,
Philadelphia, PA, 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 10 through 23. 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. IMS HEALTH conducts a number of different ongoing surveys of drug use, which it then sells to pharmaceutical manufacturers and to government agencies. Perhaps the most useful of these for pharmacoepidemiology 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, that is it contains relatively few individuals, relative to the number of individuals that are being generalized to and the number of variables that are being studied. Nevertheless, the results obtained using the NDTI tend to be relatively stable over time,¹ and agree fairly well with other sources of drug utilization information.² NDTI has proven itself very useful, and it has been used often for pharmacoepidemiology research.

A few of the other IMS HEALTH databases include the following: (i) the National Prescription Audit Plus™ (NPA™), 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 HEALTH electronically; (ii) LifeLink, a patient longitudinal database of medical and pharmacy claims; (iii) Retail Perspective™ and Provider Perspective™, two audits that provide national sales dollar estimates of pharmaceutical products purchased by retail drug stores, food stores, mass merchandise, closed wall HMOs, nonfederal hospitals, federal facilities, clinics and long-term care facilities. The Retail Perspective Report, based on data from both audits, covers over 90% of drug sales. The existence of these and other IMS HEALTH databases changes over time, as new products are developed and old products are discontinued. These and other IMS HEALTH databases can sometimes be useful for pharmacoepidemiology research, but the NDTI is generally the most useful. Generously, IMS HEALTH 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, as well, 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 is published in an annual publication entitled *Highlights of Drug Utilization in Office Practice*.³ Other special analyses can be requested, as well. Other potential sources of drug utilization data include all of the databases described in Chapters 12 through 23.

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, provided pharmacy services for the entire country.⁴ 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 29).

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 analytic studies. In some drug utilization datasets, 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 datasets information *is* available on diagnoses, but it is inappropriate. For example, the diagnosis data in NDTI includes the indication for treatment rather than the results of treatment. Finally, in some drug utilization datasets the information on diagnoses 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 29.

DISEASE INCIDENCE DATA

Data Sources

The major source of disease incidence information useful for this type of study is 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.⁵ 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.⁶ 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 statistic that are recorded by most countries include birth data, marriage data, divorce data, and so on. These are less likely to be useful for pharmacoepidemiology 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.^{6,7} These can be used to calculate incidence rates of cancer. For example, a 1976 study demonstrated that the increased use of postmenopausal estrogens was

accompanied by a sharp increase in the incidence of endometrial cancer, the latter evaluated using data from cancer registries.⁷ 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 area⁸ (see Chapter 42).

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, MA, which has been the source of an enormous amount of knowledge about risk factors for cardiovascular disease, as well as other diseases.⁹

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 and the Institutional Population Survey investigate 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.¹⁰ 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 problem 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.¹¹ 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.¹² As lung scanning procedures have become widely available, one would expect that a larger proportion of patients with pulmonary embolism may be receiving 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 methyldopa and biliary carcinoma using data from multiple international cancer registries, all but one showed no association between drug sales and disease incidence or mortality.¹³ 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 pattern 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.¹⁴

Fifth, there can be changes in population demographics. The aging of the US population now under way 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.¹⁵ 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 which 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 10 and 11). 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.¹⁶

OLMSTEAD COUNTY MEDICAL RECORDS

Another approach to performing case–control studies takes advantage of the unique medical record system in Olmstead County, MN.¹⁷ The Mayo Clinic and its affiliated clinics and hospitals provide the medical care for most of the 250 000 residents of Olmstead County. These data have been supplemented since 1966 by information on diagnostic and surgical procedures from the other medical groups and hospitals in Olmstead County, as well as the few independent practitioners who were not part of the Mayo Clinic system. 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 diagnosis files for its clients, as a

by-product of its provincial health plan.¹⁸ As with the Mayo Clinic data, computerized prescription drug information is generally unavailable and, so, needs to be obtained on an *ad hoc* basis.

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 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 15 through 23; 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 over-matching; 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 which 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.

The validity of exposure data collected from patients as part of *ad hoc* case–control studies is discussed in Chapter 39.

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.^{19–21} Overall, the advantages and disadvantages of this approach are identical to those of case–control surveillance, described in Chapter 13. 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, which allows 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. 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 39.

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 the pharmacist, asking him 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.

Until recently, the major pioneer in the use of pharmacy-based methods of pharmacoepidemiology had been the pharmacoepidemiology group at Upjohn Pharmaceuticals. The results of a feasibility study using this approach have been published.²² Briefly, the study consisted of the identification and followup 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 followup 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.²³ 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.^{24,25} Patients filling a prescription for a target medication are presented an announcement of the study along with their prescription. Patients who agree to participate then are asked to report during the next month 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.^{24,25} 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 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, sufficiently severe to mask any real result due to nondifferential misclassification.

Finally, more recently, pharmacy-based surveillance was used to conduct a massive study of parenteral ketorolac.^{26–28} A retrospective cohort study between 18 November 1991 and 31 August 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 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 GI bleeding, especially in the elderly, with higher doses, and with use over 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 felt to be similar, but that improving the use of ketorolac (e.g., duration < 5 days) would improve the risk/benefit balance further, and 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 they protected the drug's availability in some markets where concerns had been raised.

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 followup 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 better quantitate the incidence of first dose syncope, which was a well recognized adverse effect of this drug.²⁹ 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 H-2 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.³⁰

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 a million dollars each, without taking into account the considerable time of the pharmaceutical representatives.³¹

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.³²

It often would be difficult or impossible to enroll appropriate controls for a new drug in this type of study. For example, no other H-2 blocker was on the market at the time cimetidine was marketed. When a United Kingdom postmarketing surveillance study was performed which 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 the cimetidine.³³ 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 which 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, while in fact they are market seeding studies. This practice is unfortunate, however, and should be abandoned. In addition to being of questionable honesty, 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.³⁴

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 15 to 23).

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 records for any intervening care received by the patient.³⁵

Finally, other approaches can be thought of as well, such as by systematically approaching 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 approach 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.³⁶ 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.³⁷ 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.³⁸ 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 Health Maintenance Organization randomly allocated its pharmacies to dispense either the wax-matrix formulation or the microencapsulated formulation.

As another example, we have planned a large scale double-blind randomized clinical trial of two different drugs in the same drug class. 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 33, on the use of randomized clinical trials for pharmacoepidemiology studies.

ADDITIONAL DATABASES USEFUL FOR PHARMACOEPIDEMOLOGY RESEARCH

Finally, there are a number of other databases potentially useful for pharmacoepidemiology 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.³⁹ The particular advantages of this system is that it collects nationwide data and some of the notifications are mandatory. The particular disadvantages are the small population size (Finland's population is only 4.6 million, in total, and there are only 60 000 individuals per year with data reported to the Register of Persons Entitled to Free Drugs), and the limited staff available to make use of the data set. Nevertheless, it has been used on occasion for formal analytic research in pharmacoepidemiology. For example, a paper reported a case-control study that did not confirm the initial reports of an association between reserpine and breast cancer.⁴⁰

An analogous database was created using the Oxford Record Linkage Study.⁴¹ It included four group practices and one solo practice, a total of 16 general practitioners' practices, 33 000 patients at any point in time, and 43 117 patients over the two-year time period of study. Demographic information came from the computerized Oxford Community Health Project. Medical outcome data were collected by obtaining photocopies of prescriptions from the Prescription Pricing Authority, which were then computerized.

Another database occasionally being used for pharmacoepidemiology research is the longstanding Regenstrief Medical Record System.⁴² This database contains all laboratory, pharmacy, and appointment information for a network of inner city facilities in Indianapolis, including a 340-bed hospital and associated outpatient practices. It has been used for many studies, although only a few pharmacoepidemiology studies (e.g., 43-45). While 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 pharmacoepidemiology research, very different from Regenstrief, is IMS America's MediPlus database. This is a UK medical record database, very similar to the General Practice Research Database described in detail in Chapter 23. 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. There have been relatively few papers published to date using this resource.⁴⁶⁻⁵⁰ An analogous database exists in Germany.⁵¹

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, UT (540 beds).⁵² This is a computerized hospital information system designed to provide administrative, financial, and clinical hospital services. This system consists of a large database capable of assisting with the management of clinical data, order communications, results review, the production of comprehensive clinical reports, and decision making. It has taken over 20 years to develop the current system, and it is continuously updated and changed, and new functions are added. There is patient-based information on the drugs administered, which can be linked to events, other therapies, and procedures. To date, the epidemiologic applications of the HELP system have been primarily in the area of infectious diseases and antibiotic use (e.g., 53, 54). 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.⁵⁵

A number of new databases are emerging from Italy, as well. There is one in particular, from the Italian region of Friuli-Venezia Giulia that has been productive of a number of papers in the international literature.⁵⁶⁻⁵⁸ Of the 1.2 million inhabitants of that region, this database contains information on those who are covered by the Italian National Health Service, including outpatient prescriptions, demographic information, and hospital services provided. A methodological paper from that database indicates that the diagnoses of upper gastrointestinal bleeding and perforation are sufficiently valid to be useful for pharmacoepidemiology research.⁵⁸ This is a database that is likely to be used much more often for pharmacoepidemiology research in the next few years.

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.

REFERENCES

1. IMS Research Group. *A Special Analysis of the National Disease and Therapeutics Index Service for the Pharmaceutical Industry*. Ambler, PA: IMS America, 1979.
2. Knapp D. *Comparisons of NAMCS, NDTI, and Medicaid Data*. Washington: US Food and Drug Administration, 1983.

3. *Highlights of Drug Utilization in Office Practice—National Ambulatory Medical Care Survey, 1985*, Advance Data Number 137. Hyattsville, MD: National Center for Health Statistics, 1987.
4. Lilja J. The nationalization of the Swedish pharmacies. *Soc Sci Med* 1987; **24**: 423–9.
5. Edlavitch SA, Feinleib M, Anello C. A potential use of the National Death Index for postmarketing drug surveillance. *J Am Med Assoc* 1985; **253**: 1292–5.
6. Smith ME, Newcombe HB. Automated follow-up facilities in Canada for monitoring delayed health effects. *Am J Public Health* 1980; **70**: 1261–8.
7. Weiss NS, Szekely DR, Austin DF. Increasing incidence of endometrial cancer in the United States. *New Engl J Med* 1976; **294**: 1259–62.
8. Oakley GP. Population and case-control surveillance in the search for environmental causes of birth defects. *Public Health Rep* 1984; **99**: 465–8.
9. Dawber TR. The Framingham study. *The Epidemiology of Atherosclerotic Disease*. Cambridge, Harvard University Press, 1980.
10. Stolley PD. The use of vital and morbidity statistics for the detection of adverse drug reactions and for monitoring of drug safety. *J Clin Pharmacol* 1982; **22**: 499–504.
11. Dalen JE, Alpert JS. National history of pulmonary embolism. *Prog Cardiovasc Dis* 1975; **517**: 259–70.
12. Hull RD, Raskob GE, Hirsh J. The diagnosis of clinically suspected pulmonary embolism. *Chest* 1986; **89**: 417S–25S.
13. Strom BL, Hibberd PL, Stolley PD. No evidence of association between methyldopa and biliary carcinoma. *Int J Epidemiol* 1985; **14**: 86–90.
14. Stolley PD, Strom BL. Evaluating and monitoring the safety and efficacy of drug therapy and surgery. *J Chron Dis* 1986; **39**: 1145–55.
15. Gillum RF, Folsom AR, Blackburn H. Decline in coronary heart disease mortality. Old questions and new facts. *Am J Med* 1984; **76**: 1055–65.
16. Strom BL, West SL, Sim E, Carson JL. Epidemiology of the acute flank pain syndrome from suprofen. *Clin Pharmacol Ther* 1989; **46**: 693–9.
17. Kurland LT, Molgaard CA. The patient record in epidemiology. *Sci Am* 1981; **245**: 54–63.
18. Roos LL, Nicol JP. Building individual histories with registries. *Med Care* 1983; **21**: 955–69.
19. Schlesselman JJ. *Case-Control Studies—Design, Conduct, Analysis*. New York: Oxford University Press, 1982.
20. Ibrahim MA (ed.). Case-control study: consensus and controversy. *J Chron Dis* 1979; **32**: 1.
21. Breslow NE, Day NE. *The Analysis of Case-Control Studies* (Statistical Methods in Cancer Research 1). Lyon: International Agency for Research on Cancer, 1980.
22. Borden EK, Lee JG. A methodologic study of postmarketing drug evaluation using a pharmacy-based approach. *J Chron Dis* 1982; **35**: 803–16.
23. Luscombe FA. Methodologic issues in pharmacy-based post-marketing surveillance. *Drug Inf J* 1985; **19**: 269–74.
24. Fisher S, Bryant SG. Postmarketing surveillance of adverse drug reactions: patient self-monitoring. *J Am Board Fam Pract* 1992; **5**: 17–25.
25. Fisher S, Bryant SG, Kent TA. Postmarketing surveillance by patient self-monitoring: trazodone versus fluoxetine. *J Clin Psychopharmacol* 1993; **13**: 235–42.
26. Strom BL, Berlin JA, Kinman JL, Spitz PW, Hennessy S, Feldman H, Kimmel S, Carson JL. Parenteral ketorolac and risk of gastrointestinal and operative site bleeding: a postmarketing surveillance study. *J Am Med Assoc* 1996; **275**: 376–82.
27. Feldman HI, Kinman JL, Berlin JA, Hennessy S, Kimmel SE, Farrar J, Carson JL, Strom BL. Parenteral ketorolac: the risk for acute renal failure. *Ann Intern Med* 1997; **126**: 193–9.
28. Hennessy S, Kinman JL, Berlin JA, Feldman HI, Carson JL, Kimmel SE, Farrar J, Harb G, Strom BL. Lack of hepatotoxic effects of parenteral ketorolac in the hospital setting. *Arch Intern Med* 1997; **157**: 2510–14.
29. Graham RM, Thornell IR, Gain JM, Bagnoli C, Oates HF, Stokes GS. Prazosin: the first dose phenomenon. *Br Med J* 1976; **2**: 1293–4.
30. Gifford LM, Aeugle ME, Myerson RM, Tannenbaum PJ. Cimetidine postmarket outpatient surveillance program. *J Am Med Assoc* 1980; **243**: 1532–5.
31. Joint Commission on Prescription Drug Use. *Final Report*. Washington: Joint Commission on Prescription Drug Use, 1980.
32. Rossi AC, Knapp DE, Anello C, O'Neill RT, Graham CF, Mendelis PS, Stanley GR. Discovery of adverse drug reactions. A comparison of selected phase IV studies with spontaneous reporting methods. *J Am Med Assoc* 1983; **249**: 2226–8.
33. Colin-Jones DG, Langman MJS, Lawson DH, Vessey MP. Postmarketing surveillance of the safety of cimetidine: 12 month mortality report. *Br Med J* 1983; **286**: 1713–6.
34. 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.
35. Colin-Jones DG, Langman MJS, Lawson DH, Vessey MP. Cimetidine and gastric cancer: preliminary report from postmarketing surveillance study. *Br Med J* 1982; **285**: 1311–3.
36. Bell RL, Smith EO. Clinical trials in postmarketing surveillance of drugs. *Control Clin Trials* 1982; **3**: 61–7.

37. 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.
38. Bradford RH, Shear CL, Chremos AN, Dujovne C, Downton M, Franklin FA, Gould AL, Hesney M, Higgins J, Hurley DP. Expanded clinical evaluation of Lovastatin (EXCEL) study results. *Arch Intern Med* 1991; **151**: 43–9.
39. Idanpaan-Heikkila J. Finland. In: Inman WHW, ed., *Monitoring for Drug Safety*. Boston: MTP Press, 1986; 83.
40. Aromaa A, Hakama M, Hakulinen T, Saxen E, Teppo L, Ida lan-Heikkila 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.
41. Skegg DCG, Doll R. Record linkage for drug monitoring. *J Epidemiol Community Health* 1981; **35**: 25–31.
42. 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.
43. Williams CL, Johnstone BM, Kesterson JG, Javor KA, Schmetzer AD. Evaluation of antipsychotic and concomitant medication use patterns in patients with schizophrenia. *Med Care* 1999; **37** (suppl 4 Lilly): AS81–6.
44. MacDonald CJ, Tierney WM, Overhage JM, Martin DK, Smith B, Wodniak C, Blevins L, Warvel J, Meeks-Johnson J. The Regenstrief Medical Record System—experience with MD order entry and community-wide extensions. *Proceedings—the Annual Symposium on Computer Applications in Medical Care* 1994; 1059.
45. Tierney WM, Miller ME, Overhage JM, McDonald CJ. Physician inpatient order writing on microcomputer workstations. Effects on resource utilization. *J Am Med Assoc* 1993; **269**: 379–83.
46. Lloyd DC, Harris CM, Roberts DJ. Specific therapeutic group age–sex related prescribing units (STAR-PU): weightings for analysing general practices' prescribing in England. *Br Med J* 1996; **311**: 991–4.
47. Harris CM, Dajda R. The scale of repeat prescribing. *Br J Gen Pract* 1996; **46**: 649–53.
48. Hughes D, McGuire A. The direct costs to the NHS of discontinuing and switching prescriptions for hypertension. *J Hum Hypertens* 1998; **12**: 533–7.
49. Eggleston A, Wigerinck A, Huijghebaert S, Dubois D, Haycox A. Cost effectiveness of treatment for gastro-oesophageal reflux disease in clinical practice: a clinical database analysis. *Gut* 1998; **42**: 13–6.
50. Lawrenson R, Farmer RD, Williams TJ, Leydon GM. Venous thromboembolic disease and combined oral contraceptives: a re-analysis of the MediPlus database. *Hum Reprod* 1999; **14**: 1500–5.
51. Farmer RD, Todd JC, Lewis MA, MacRae KD, Williams TJ. The risks of venous thromboembolic disease among German women using oral contraceptives: a database study. *Contraception* 1998; **57**: 67–70.
52. Classen DC, Burke JP. The computer-based patient record: the role of the hospital epidemiologist. *Infect Control Hosp Epidemiol* 1995; **16**: 729–36.
53. Evans RS, Pestotnik SL, Classen DC, Clemmer TP, Weaver LK, Orme JF Jr, Lloyd JF, Burke JP. A computer-assisted management program for antibiotics and other anti-infective agents. *New Engl J Med* 1998; **338**: 259–60.
54. Classen DC, Burke JP, Pestotnik SL, Lloyd JF. Clinical and financial impact of intravenous erythromycin therapy in hospitalized patients. *Ann Pharmacother* 1999; **33**: 669–73.
55. Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. *J Am Med Assoc* 1991; **266**: 2847–51.
56. 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.
57. Garcia Rodriguez LA, Cattaruzzi C, Troncon MG, Agostinis L. Risk of hospitalization for upper gastrointestinal tract bleeding associated with ketorolac, other nonsteroidal anti-inflammatory drugs, calcium antagonists, and other antihypertensive drugs. *Arch Intern Med* 1998; **158**: 33–9.
58. Cattaruzzi C, Troncon MG, Agostinis L, Garcia Rodriguez LA. Positive predictive value of ICD-9th codes for upper gastrointestinal bleeding and perforation in the Sistema Informativo Sanitario Regionale database. *J Clin Epidemiol* 1999; **52**: 499–502.

25

How Should One Perform Pharmacoepidemiology Studies? Choosing Among the Available Alternatives

BRIAN L. STROM

Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine,
Philadelphia, PA, 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 10 through 24 presented each of the different approaches that have been developed to perform pharmacoepidemiology studies efficiently, despite 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 PHARMACOEPIDEMIOLOGY STUDIES

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. 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, this author feels strongly that it is important to tailor the choice of pharmacoepidemiology resource to the question to be addressed. One may want to use more than one approach, 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.

Table 25.1. Comparative characteristics of pharmacoepidemiology data resources^a

Pharmacoepidemiology approach	Relative size	Relative cost	Relative speed	Representativeness	Population-based	Cohort studies possible	Case-control studies possible
Spontaneous reporting	+++	+	+++	+++	-	-	+
Intensive hospital-based cohort studies	+	+++	+	variable	-	+	-
Group Health Cooperative of Puget Sound	+	+	+++	+++	+	+	+
Kaiser Permanente Medical Care Programs	++	+	+++	+++	+	+	+
Harvard Pilgrim Health Care	+	+	+++	+++	+	+	+
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	+++	-	-	-	+	-

^a See the text of this chapter for descriptions of the column headings, and previous chapters for descriptions of the data resources.

Table 25.1. *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 followup
Spontaneous reporting	+++	++	-	+	+++	fall 1969 to date	N/A
Intensive hospital-based cohort studies	++++	++++	+++	+	-	1966-1976	N/A
Group Health Cooperative of Puget Sound	++	+ to +++	+++	+	-	1972 to date	15%/year × 2 years, then 5%/year 3%/year after 2 years
Kaiser Permanente Medical Care Programs	++	+ to +++	+++	-	-	- ^b	14%/year Unknown 20-25%/year nil
Harvard Pilgrim Health Care	++	+ to +++	++	+	+++	4/88 to date	
UnitedHealth Group	+	+ to +++	++	-	++	1990 to date	
Medicaid databases	+++	+ to +++	++	-	++	- ^c	
Health databases in Saskatchewan	+++	+ to +++	++	-	+	9/75-6/87, 1/89 to date	
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	N/A
Pharmacy-based surveillance: outpatient	+++	+++	+++	-	+	as desired	25% variable
<i>Ad hoc</i> cohort studies	+++	+++	+++	+	+++	as desired	variable
Pharmacoepidemiology randomized trials	+++	+++	++++	+	++++	as desired	variable

^bIn the Northern California Program, computerized pharmacy data are complete for: (i) 1969-1973 for one large pharmacy; (ii) 1981 to date, for a 1% sample, and (iii) March 1984 through at least April 1985 for one large pharmacy, all pharmacies since 1986 in Oregon, and since 1994 in Northern California. Discharge diagnoses are computerized from 1971 to date.

^cTennessee: 1973 to date; other programs 1980 to date.

Table 25.2. Characteristics of research questions and their impact on the choice of pharmacoepidemiology data resource^a

Pharmacoepidemiology approach	Hypothesis generating ^b	Hypothesis strengthening ^c	Hypothesis testing ^d	Study of benefits (vs. risks)	Incidence rates desired	Low incidence outcome
Spontaneous reporting	++++	+	-	+	-	++++
Intensive hospital-based cohort studies	+	++	+++	+	++++	+
Group Health Cooperative of Puget Sound	+	++++	+++	+	+++	+
Kaiser Permanente Medical Care Programs	++	+++	+++	+	+++	+++
Harvard Pilgrim Health Care	+	++++	+++	+	+++	+
UnitedHealth Group	+	++++	+++	+	+++	+++
Medicaid databases	++	++++	+	+	+++	+++
Health databases in Saskatchewan	+	++++	+++	+	+++	++
Netherlands	+	++	++	+	+++	++++
Tayside	+	++++	++	+	+++	+
GPRD	+	++++	+++	+	+++	++
Ad hoc analyses of secular trends	+	+++	-	+	-	++++
Case-control surveillance	++	+++	+++	+	-	+++
Registry-based case-control studies	+	+++	++	+	+	+++
Ad hoc case-control studies	+	++	+++	+	+	+++
Prescription Event Monitoring	+	++	+++	+	++	++++
Pharmacy-based surveillance: outpatient	+	++	+++	+	++	++
Ad hoc cohort studies	+	++	+++	+	++	++
Pharmacoepidemiology randomized trials	+	+	++++	++++	+++	+

^aSee the text of this chapter for descriptions of the column headings, and previous chapters for descriptions of the data resources.

^bHypothesis-generating studies are studies designed to raise new questions about possible unexpected drug effects, whether adverse or beneficial.

^cHypothesis-strengthening studies are studies designed to provide support for, although not definitive evidence for, existing hypotheses.

^dHypothesis-testing studies are studies designed to evaluate in detail hypotheses raised elsewhere.

Table 25.2. *continued*

Pharmacoepidemiology approach	Low prevalence exposure	Important confounders	Drug use inpatient (vs. 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	++++	+	++++	+++	+	+	+++	+++
Intensive hospital-based cohort studies	+	++++	++++	-	+++	-	+	++
Group Health Cooperative of Puget Sound	+	+++	-	-	+++	+++	++	+++
Kaiser Permanente Medical Care Programs	+++	+++	-	+	-	++	++	+++
Harvard Pilgrim Health Care	-	-	+++	+++	-	-	++	+++
UnitedHealth Group	+++	++	-	+++	-	-	++	+++
Medicaid databases	++++	+	-	+++	-	+	+++	+++
Health databases in Saskatchewan	++	++	-	++	-	+++	++	+++
Netherlands	++++	+	-	-	-	++	+++	+++
Tayside	+	+	-	-	-	++	+++	+++
GPRD	++	++	+/-	+++	-	-	++	+++
Ad hoc analyses of secular trends	-	-	-	-	-	-	+	++
Case-control surveillance	+	+++	-	-	-	++	+	to +++
Registry-based case-control studies	+	++	-	+	+	++	+	+
Ad hoc case-control studies	+	+++	+++	+++	-	++	+	+
Prescription Event Monitoring	++++	++	-	+++	++	+	+++	to +++
Pharmacy-based surveillance: outpatient	+++	+++	-	+++	++	+	+++	+
Ad hoc cohort studies	+++	+++	+	+++	++	+	+++	+
Pharmacoepidemiology randomized trials	+++	++++	+	+++	+++	+	+++	+

Regardless, investigators are often left with a difficult and complex choice.

In order to explain how to choose among the available pharmacoepidemiology 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 25.1). 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 2). The weightings 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 feelings of this author alone. 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 PHARMACOEPIDEMOLOGY DATA RESOURCES

Table 25.1 lists each of the different pharmacoepidemiology 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 Record Linkage System, and Prescription Event Monitoring cover entire countries or large fractions thereof. Medicaid databases are next largest, with UnitedHealth Group approaching that, with about 13 million persons. The GPRD database 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 400 000. 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 followup 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 to *relative speed*, 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. Spontaneous reporting systems, Prescription Event Monitoring, the health databases in Saskatchewan, the Netherlands System, and the Tayside system each include entire countries, provinces, or states and, so, are typical populations. 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. Group Health Cooperative, Harvard Pilgrim Health Care, Kaiser, and UnitedHealth include

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 techniques are listed as “variable”, meaning their representativeness depends on which hospitals are recruited into the study. *Ad hoc* studies are listed as “as desired,” as 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 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,¹ but this is unusual (see also Chapters 2 and 24).

Whether cohort studies are possible 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 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 except the Netherlands system, which lacks access to linked diagnosis data, as well as any of the systems designed to perform case-control studies. Even the Netherlands is developing that capability, using PHARMO. Data from spontaneous reporting systems can be used for case finding for case-control studies, although this has been done infrequently.²

The *validity of the exposure data* is most certain in hospital-based settings, in which one can be reasonably certain of both the identity of a drug and that the patient ingested it. Exposure data in spontaneous reporting systems originate 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. They 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 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.³ However, this would be expected to vary, depending on the type of drug taken, the questioning technique used, etc. (see Chapter 39).³⁻¹¹

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. GPRD analyzes the actual medical record, rather than claims, and can access additional questionnaire data from the general practitioner, as well. Medicaid databases are no longer able to access medical records to validate their outcomes, for reasons of confidentiality.

Control of confounding refers to the ability to control for confounding variables. The most powerful approach to controlling 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. Techniques that collect sufficient information to control for known and measurable variables are next most effective. These include intensive hospital-based cohort studies, 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, and Harvard Pilgrim Health Care 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 able to access medical records. Finally, spontaneous reporting systems and analyses of trends have no control of confounding.

Relatively few of the data systems have data on *inpatient drug use*. The exceptions include spontaneous reporting systems, intensive hospital based cohort studies, Group Health (only since 1989, and not yet used for research), Harvard Pilgrim Health Care, *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, GPRD, Harvard Pilgrim Health Care, 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 *dates of the available data* differ substantially among the different resources, as does the degree of loss to followup. They are specified in Table 25.1.

CHARACTERISTICS OF RESEARCH QUESTIONS AND THEIR IMPACT ON THE CHOICE OF PHARMACOEPIDEMOLOGY 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 25.2).

Pharmacoepidemiology studies can be mounted 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 and approaches can and do raise such questions, through incidental findings in studies mounted for other reasons. In addition, virtually any case-control

study could be used, in principle, to screen for possible drug causes of the disease under study, and virtually any cohort study could be used to screen for unexpected outcomes from the 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 dataset, 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 this was not an *a priori* hypothesis, a hypothesis-testing type of study could 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, although some limited additional work on this is under way. 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 a computerized form, both exposure data and outcome data are most useful. Those that can rapidly access only one of these, exposure or outcome data, are next most useful, while those that need to gather both are least useful, because of the time and expense that would be entailed.

Hypothesis-testing studies are studies designed to evaluate in detail hypotheses raised elsewhere. Such studies must be able to have simultaneous comparison groups and must be able to control for most known potential confounding variables. For these reasons, spontaneous reporting systems cannot be used for this purpose, as they cannot be used to conduct studies with simultaneous controls (with rare exception—see reference 2). 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 which allow access to primary records but not the patient are next most useful.

The research implications of questions about the *beneficial effects* of drugs are different, depending upon whether the beneficial effects of interest are expected or unexpected 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 cases one is studying side-effects. Studies of expected beneficial effects, or drug efficacy, raise the special methodologic problem of confounding by the indication: patients who receive a drug are different from those who do not in a way which usually is related to the outcome under investigation in the study. This issue is discussed in detail in Chapter 34. 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, like Prescription Event Monitoring, *ad hoc* cohort studies, and outpatient pharmacy-based cohort studies. On the other hand, these approaches are 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 either 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 or Medicaid databases, which cover a large proportion of the United States, or GPRD. Other 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 techniques. Because they recruit patients on the basis of the patients suffering from a disease, they are more efficient than attempting to perform such studies using analogous cohort studies. Kaiser in Northern California includes 2.8 million subscribers, Kaiser in Southern California over 2 million, and Kaiser Northwest about 400 000. 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. Group Health Cooperative includes fewer individuals and so would be less useful to answer questions about uncommon outcomes. Harvard Pilgrim Health Care and Tayside are also small. Finally, intensive hospital-based cohort studies and 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 either spontaneous reporting, the Netherlands system, or Prescription Event Monitoring, which cover entire countries. Alternatively, one could use UnitedHealth Group or Medicaid databases, which cover a large proportion of the United States, or GPRD. 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 control exposure, could assure an adequate number of exposed individuals. Other 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 *drug use in question is inpatient* drug use, then the data resource must obviously be capable of collecting data on inpatient drug exposures. The number of approaches which have this capability are limited, and include spontaneous reporting systems, intensive hospital-based cohort studies, Harvard Pilgrim Health Care, and inpatient pharmacy-based surveillance systems. *Ad hoc* studies could also, of course, be designed to collect such information.

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 GPRD, which collect their data from general practitioners, are excellent sources of data for this type of question. Harvard Pilgrim Health Care is similar, with data collected in an automated fashion. 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 are 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. Only randomized trials can be specifically designed to be certain this information is collected. Intensive hospital-based cohort studies might observe such outcomes, but are unlikely to do so unless data gathering was specifically targeted to observe the outcome in question. *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 and Kaiser Permanente have even longer followup time available. However, as HMOs they suffer from some turnover, albeit more modest after the first few years. 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 techniques 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. Finally, one cannot study long-term drug effects using inpatient approaches.

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 which can collect a significant number of such exposures quickly. Any of the cohort study approaches 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 next most useful, especially Medicaid databases, as their large population base will allow one to accumulate a sufficient number of exposed individuals rapidly, so one can perform a study sooner. In some cases, there is a delay until the drug is available on the program's formulary, however. Intensive hospital-based cohort studies will take a long time before an adequate sample size is collected, although this may be the only viable approach to studying a drug used in the hospital only. *Ad hoc* analyses of secular trends and case-control studies, by whatever approach, must wait until sufficient drug exposure has occurred that it can affect the outcome variable being studied.

Finally, if *one needs an answer to a question urgently*, potentially the fastest approach, if the 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. Also very useful are the Medicaid databases. In many states, the drugs are available almost immediately; both exposure and diagnosis data are available in a computerized format; both inpatient and outpatient diagnosis data are available in a computerized format, if desired; the time delay between the exposure or the medical event and its appearance in the computerized record is relatively short; and the large population base allows one to attain the needed sample sizes quickly. The other computerized databases are also very useful for these purposes. 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 techniques 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 nonsteroidal anti-inflammatory drugs (NSAIDs) cause upper gastrointestinal bleeding and, if so, how often. One could examine the manufacturer's premarketing data from clinical trials, but the number of patients included is not likely to be large enough to study clinical bleeding, and the setting is very artificial. Alternatively, one could examine premarketing studies using more sensitive outcome measures, such as endoscopy. However,

these are even more artificial. Instead, one could use any of the databases to address the question quickly, as they have data on drug exposures that preceded the hospital admission. Some could only investigate gastrointestinal bleeding resulting in hospitalization (e.g., Kaiser Permanente, except via chart review, or Tayside). Others could 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, that 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 a signal generated by the spontaneous reporting system, designing a study to investigate whether a drug which has been on the market for, say, five years is a cause of a relatively rare condition, such as allergic hypersensitivity reactions. Because of the infrequency of the disease, one would need to draw on a very large population. The best alternatives would be Medicaid databases, *ad hoc* analyses of trends, case-control studies, or Prescription Event Monitoring. If this is a drug used primarily as an inpatient, then the Harvard Pilgrim Health Care would be the only viable choice, if the outcome were common enough to be detectable with their sample sizes. If this was a drug used primarily as an outpatient, then the other techniques would be feasible. 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.

CONCLUSIONS

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.

REFERENCES

1. 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. *J Am Med Assoc* 1986; **256**: 1749-57.
2. Strom BL, West SL, Sim E, Carson JL. Epidemiology of the acute flank pain syndrome from suprofen. *Clin Pharmacol Ther* 1989; **46**: 693-9.
3. Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformations. *Am J Public Health* 1967; **57**: 2071-5.
4. Glass R, Johnson B, Vessey M. Accuracy of recall of histories of oral contraceptive use. *Br J Prev Soc Med* 1974; **28**: 273-5.
5. Stolley PD, Tonascia JA, Sartwell PE *et al*. Agreement rates between oral contraceptive users and prescribers in relation to drug use histories. *Am J Epidemiol* 1978; **107**: 226-35.
6. Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 1982; **116**: 114-22.
7. Rosenberg MJ, Layde PM, Ory HW *et al*. Agreement between women's histories of oral contraceptive use and physician records. *Int J Epidemiol* 1983; **12**: 84-7.
8. Schwarz A, Faber U, Borner K *et al*. Reliability of drug history in analgesic users. *Lancet* 1984; **2**: 1163-4.
9. Coulter A, Vessey M, McPherson K. The ability of women to recall their oral contraceptive histories. *Contraception* 1986; **33**: 127-39.
10. Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670-6.
11. Persson I, Bergkvist L, Adami HO. Reliability of women's histories of climacteric oestrogen treatment assessed by prescription forms. *Int J Epidemiol* 1987; **16**: 222-8.

Part IV

SELECTED SPECIAL APPLICATIONS
AND METHODOLOGICAL ISSUES IN
PHARMACOEPIDEMIOLOGY

26

Bioethical Issues in Pharmacoepidemiology Research

DAVID CASARETT AND JASON KARLAWISH

Department of Medicine, Division of Geriatrics, and Center for Bioethics, University of Pennsylvania, Philadelphia, PA, USA

ELIZABETH ANDREWS

Glaxo Wellcome, Worldwide Epidemiology, Research Triangle Park, NC, USA

ARTHUR CAPLAN

Center for Bioethics, University of Pennsylvania, Philadelphia, PA, USA

INTRODUCTION

In the last 50 years, as medical research has evolved rapidly, 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.¹⁻³ As a result, research ethics has focused primarily on protecting human subjects from the risks of research. The goal has been to minimize risks to subjects, rather than minimizing the risks and maximizing the potential benefits for both subjects and society.⁴ Themes that run through many of these scandals are scientists' failure to adequately review and disclose research risks and potential benefits, 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.

Research ethics and research practice have become separate and even, sometimes, antagonistic enterprises. The role society expects of ethics is to regulate science. Scientific practice reflects this fact. IRB review and the practice of informed consent have become as integral to the design of a clinical research as sample size calculations, the accurate measurement of endpoints, or robust statistical analysis.

These and other requirements have been remarkably effective in defining the limits of ethical research, and have made it much less likely that the most egregious ethical errors of the past will be repeated. Overall, they should be viewed as welcome additions to the practice of clinical research. However, serious scientific and ethical problems may arise when the requirements that were developed to guide clinical research are applied to other kinds of research. In particular, standard protections in clinical research are not

easily exported and applied to the very different challenges of epidemiologic research. Therefore, as these rules have been applied to pharmacoepidemiologic research, the result has been the parallel development of modifications to the ethical guidelines and principles, on one hand, and increasing consternation and confusion, on the other, about how these modifications should be applied.

The central problem has been that, while the ethics of human subjects research has been built upon the protection of human subjects, the human subjects involved in pharmacoepidemiologic research are quite different. Indeed, it may be difficult to see why and how the analysis of a dataset makes the patients whose information contributes to that set “human subjects” and the research in need of ethics review board review. 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 is not the issue in pharmacoepidemiology research. It is almost always what in law and philosophy is referred to as “wrong”, that is—a violation of a person’s rights. 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.⁵⁻⁷ Most of these guidelines apply equally well to pharmacoepidemiology research, although this field has begun to develop its own principles.⁷ These guidelines and regulations have made it clear that the protection of subjects in epidemiology research represents only one part of the ethical obligations of epidemiology investigators. 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.⁵⁻⁷

Although these guidelines acknowledge a range of ethical obligations, one of these, the investigators’ obligations to subjects, has clearly proven to be the most challenging. This is because the procedures of ethical research, like ethics board review and informed consent, may be overly protectionistic or prohibitively difficult in epidemiologic research. Ethical concerns about pharmacoepidemiology research, and more broadly about epidemiology research, have therefore focused on the kinds of research that require ethics board review and the kinds of research that require the subject’s informed consent.

The answers to these questions define the ethical procedures that allow researchers to have access to information gathered for clinical and administrative purposes. Therefore, investigators face a considerable challenge. They must protect patients’ privacy and confidentiality in a way that accomplishes research goals accurately and efficiently. This challenge lies at the heart of the ethics of pharmacoepidemiology research.

National and international organizations have created principles that provide a backdrop to the research framework, the best established being those adopted by the Organisation for Economic Co-operation and Development in 1980.⁸ These recommendations suggest that limits to the collection of data should be sought, that the quality of data is important, that data use should be specified in advance, and that investigators should adhere to specified uses. Finally, the Co-operative suggests a requirement of “openness,” that is, a requirement that goals, uses, and access to data should be a matter of public record, and that individuals should be able to determine whether and how data about them are being used. Despite general agreement about these and other principles, the international community has failed to achieve a consensus about the proper balance of protections and research progress.

The goal of this chapter is to present an overview of this balance and specifically of the challenges that arise when the principles of research ethics are applied to issues surrounding privacy and confidentiality. In order to accomplish this goal, this chapter will tend to emphasize regulations in the United States. This is not

because these regulations can or should be generalized to other countries, but simply because at the current time international guidelines vary widely and are often contradictory.^{9,10} Therefore, although the experience of the United States is not universal, these regulations provide a frame of reference for comparison. Where instructive, however, experience from other countries is discussed as well.

This chapter begins by defining the terms that describe the procedures and requirements of ethical research. These are the normative boundaries in which pharmacoepidemiology must operate in order to maintain the public's trust. If research is to move forward, pharmacoepidemiologists must develop procedures that permit them to balance a need for scientific rigor, on one hand, with respect for ethical requirements, on the other. This chapter will discuss three such strategies and the challenges that investigators face in applying them: de-linking subject identifiers from their information and modifications to both bioethics review board review and subject informed consent requirements. This chapter concludes with a critical consideration of some of the available guidelines and regulations, and recommendations for future regulatory efforts.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

The birth and subsequent development of research and scholarship in research ethics, like any field of specialized knowledge, has constructed a language that is particularly its own. This language provides a taxonomy of ethical issues in research and is essential to this discussion because it forms the foundation of any communication and discourse between the fields of ethics and epidemiology. These terms also offer an excellent vantage point from which to examine critically the current emphasis on human subjects protections and its applicability to pharmacoepidemiologic research.

RESEARCH

Any productive analysis of the ethics of pharmacoepidemiologic research is critically dependent on a clear and precise understanding of the term "research." Given the frequency with which this term is used by ethicists, investigators, and the public, a definition would seem to be a simple matter. Unfortunately, this has been far from the truth.¹¹ Yet, perhaps the best established definition is also the oldest. In its summary statement (the Belmont Report), the US National Commission for the Protection of Human Subjects defined "research" as any activity designed to "develop or contribute to generalizable knowledge."¹² This is a definition that has been embraced by other scholars, and has become the standard by which a proposed project is assessed.¹³

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 define the patterns of medication use in a local population. It is not clear, given the definition based on "generalizable knowledge," whether this project should be construed as research, clinical care, or even as a quality improvement activity. These distinctions are important because once a project is identified as "research" investigators must meet a series of requirements designed to protect the patients who are now human subjects.

This definition is particularly problematic in pharmacoepidemiologic research, because it is hard to distinguish the routine practice of epidemiology from research. The extremes are evident. The paradigmatic *practice* of epidemiology is public health case finding and surveillance for adverse drug reactions. 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. These sorts of investigations proceed, and often 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 activity can be difficult to demarcate.

HUMAN SUBJECT

Although it is important than any discussion of research and research ethics be clear about the definition of a research subject, this definition is as elusive as the definition of research, on which it depends. Broadly, though, a useful definition comes from the United States "Common Rule," the set of Federal regulations first promulgated in 1981 that govern research ethics.¹⁴ The Common Rule defines a "research subject" as "a living individual, about whom an investigator (whether professional or student) conducting research obtains either: 1) data through intervention or interaction with the individual, or 2) identifiable private information."^{14,46,102f} For pharmacoepidemiologists, the key issue here is that the use of information that can be linked to an individual constitutes a contact between an investigator and a human subject. This is true even if the information was gathered in the past and no contact occurs between the investigator and the person. A 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. In addition, its influence is far greater because the vast majority of institutions that accept these federal funds have signed an agreement, called a Multiple Project Assurance, to abide by the Common Rule requirements in all research, regardless of the source of funding. Therefore, the Common Rule serves as *de facto* law governing research at the most productive research institutions in the US and offers a reasonable working definition. Further, even when research is performed outside of the United States, if it is done with US federal support or at an institution with a Multiple Project

Assurance then it must conform to American regulations governing research ethics.

ETHICS REVIEW BOARDS

In many countries over the last 30 years, ethics review boards have become central to the practice of research. In the American context, these are committees with at least one "community representative," appointed by institutions that receive federal monies to conduct research. In other nations, there are regional or national committees that are appointed by professional organizations or government agencies.

This requirement reflects the consensus that scientists, and science, could benefit from independent review of research protocols. This idea first appeared in the World Medical Association's Declaration of Helsinki in 1964, which requires that an independent committee review all protocols. The Declaration recommends that this committee be responsible for "consideration, comment and guidance" but does not define further the committee's authority to approve or reject protocols that it finds unacceptable.^{10,pp. 214-6} These recommendations have been taken up rapidly, and review boards have become widespread. Their authority has been clarified as well, and these committees typically have the power to review and reject all research that takes place in their institution or in which their institution's investigators are involved.

In the US, while some states have enacted legislation governing human subject research, the formal system of review has evolved primarily in a manner that links federal authority and funding. A committee is required to review all research that is funded by all federal government branches that have signed on to the Common Rule.¹⁴ Examples are the National Institutes of Health, the Food and Drug Administration, and the Agency for Health Care Policy and Research. Further, as noted above, institutions that have filed a Multiple Project Assurance have agreed to abide by the Common Rule requirements in all research, regardless of the source of funding. In most other countries, research regulations are not limited by

provisions regarding funding but, instead, apply to all research conducted in that country.

The composition of these review boards varies widely across international boundaries. However, a consistent feature is the need for inclusion of expertise from outside of the scientific community. For instance, the US regulations mandate the inclusion of at least one member who is not affiliated with the institution, and one member who may be affiliated but who represents law, ethics, or another nonscience discipline.^{14,46,107} Australian regulations mandate a committee's composition by requiring a mix of genders, and by extending the inclusion of nonscience representatives.^{10, pp. 343–54} 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.

Although review boards have become a commonplace feature on the research landscape, even under US federal guidelines, not all research requires review. Certain kinds of research can receive expedited review, that is, review by the IRB chair or a designated member of the IRB instead of the full committee, and some may be exempt. 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.^{14,46,101} For example, when investigators use data in which nothing would permit the investigator to identify the individual from which the data came, ethics board review is not required. 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.^{12,46,110} 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 pharmacoepidemiology research should require review. For instance, while the Royal College of Physicians would not require review,¹⁵ the Committee of the International

Organization of Medical Societies recommends ethics board review for all research.¹⁶

PRIVACY AND CONFIDENTIALITY

In pharmacoepidemiology research, these terms are of paramount concern. Although they are often discussed together, they are distinct concepts. It is useful to distinguish them, and to describe individually the ethical basis for requirements of each. Of these, privacy is the most basic and confidentiality is in a sense derivative.

Privacy, in the setting of research, refers to 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 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.¹⁷ This is a legal way of considering a right to privacy, but privacy has an important social function as well. Viewed in this light, a right to privacy is a precondition for social interaction and cooperation because it allows and requires a degree of trust.¹⁸

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.⁶ Individuals may choose to exercise this right with the expectation, either implicit or explicit, that no one else will have access to that information without the patient's permission. This right to limit the transfer of information, to control the secondary use of information by others, is the right to confidentiality.

Like the right to privacy, the right to confidentiality is also based on a basic right to a freedom from interference, in the sense that a right to confidentiality is not possible unless there is an underlying right to privacy. However, the right to confidentiality also engenders a responsibility on the part of the person who has information about another person. The expectation that someone will

not disclose the information to a third party creates a fiduciary relationship. That is, it creates an agreement based on a mutually understood set of goals and understandings. This means that confidentiality may be more highly specified by arrangements that may be made at the time that an individual initially grants access to information. For instance, patients may have specific concerns or expectations about ways in which the information they divulge may be used. These expectations may include transfer to a third party in either identifiable or unidentifiable form, or access to particular kinds of information within a medical record, or limits as to the period of time information may be available to others.

The fundamental issue is whether information that was gathered in a clinical setting, where rules of confidentiality apply, can be used for reasons, such as research, that were not part of the conditions of that relationship. Both the law and research regulations are ambiguous over what constitutes a substantive violation of confidentiality. Does the use of records without prior authorization constitute a violation of confidentiality? Or does it constitute a risk of a violation that depends on how those records are used, and on what is done with the information?

In general, society has not articulated clear answers to these questions, in large part because the questions engage well formed but conflicting political and philosophical views about how society should organize the exchange of information. For example, proponents of communitarianism (the perspective of a community created by voluntary association) argue that the good of the individual is inextricably tied to the public good.¹⁹ Thus, ethical dichotomies that pit individuals against society (such as the unauthorized use of a person's clinical information for research) must be resolved with attention to both personal and public goods.

However, proponents of liberalism, or a rights based individualism, disagree. From this perspective, what is right exists prior to what is good. This means that any unauthorized use of a person's information threatens to violate a fundamental right to privacy and the potential good derived from that use is not a proper condition to balance against that violation.

In most states in the US, these conflicting views exist in a perhaps deliberately unresolved tension. Laws (or the absence of laws) generally allow procedures that attempt to circumscribe the extremes of either view. Laws are silent on whether medical records can be used for research without the prior authorization of the patient, although a few states have laws that records can be used for research after IRB review.¹⁷ Many European nations have very strict protections of individual rights to privacy and confidentiality. For example Iceland and Sweden have very strict requirement of individual informed consent for the use of identifiable information. Other nations lean toward a more communitarian perspective with respect to epidemiologic research.

However, US research regulations do provide a set of conditions that permit the use of records regardless of whether the patient authorized their use for research. The key features of these conditions is that the research risks are minimal and the potential violation does not adversely affect subjects' rights and welfare.^{12,46,116} The following sections will discuss both of these key arguments.

MINIMAL RISK

Although the general goal of research is to produce knowledge that will benefit society, investigators must also minimize the risks to subjects. It is axiomatic that, as risks to subjects increase, the degree of subject protections, such as ethics review and informed consent, increases as well. The concept of minimal risk attempts to operationalize a risk threshold, above which protections should be more strict. Conversely, subject protections are relaxed if a research protocol does not exceed the level of minimal risk. Although the concept of minimal risk is relatively straightforward, and would apply to most pharmacoepidemiology protocols, its definition is problematic.

According to American 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.”^{14,46.102.i} In most situations, this concept is difficult to operationalize.²⁰ This is in large part because the definition lacks a clear standard against which to compare the research risks: the daily lives of healthy or “normal” persons, or the daily lives of persons who might be subjects of the research. In pharmacoepidemiology 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, as reflected in International Codes of Ethics and professional society statements about research ethics, recognizes that subjects, or for incompetent patients, their surrogates, are to be told about the nature of research and alternatives to participation, and to have the chance to volunteer to participate. It is not surprising, therefore, that research ethics guidelines, recommendations, and regulations have stressed the procedural requirement of a subject’s informed consent. In order for a subject’s consent to be informed, he or she must understand the research and must agree to participate voluntarily, without inducement or coercion.²¹

The regulations governing research informed consent in the US, while not universal, are illustrative of these features.^{14,46.116} The US regulations convey the feature of understanding by requiring that the investigator explain the research risks, benefits, and alternatives of research participation; the confidentiality of any information obtained; and the procedures for compensation and for contacting a person responsible for the research. Voluntariness is expressed by the requirement that investigators tell subjects that participation in the research study is voluntary, and that subjects have the right to discontinue participation at any time. In some situations, informed consent may be modified to be verbal

instead of written, or even may not need to be obtained at all. Whether informed consent must always be obtained, and in what form consent should be documented, have been the subject of vigorous debate.^{22,23}

Again, while the US guidelines are not universal, they offer a helpful perspective on the complexities that this issue raises. The Common Rule requires that written informed consent be obtained in most research situations.^{14,46.116} However, it makes two notable exceptions. First, written documentation of informed consent is not required if the principal risk of the research is a breach of confidentiality and if the written record is the only link between personal data and the subjects’ identity.^{14,46.117.c} In this case, whether or not a written informed consent document is used depends on each subject’s preferences regarding linking information. Second, informed consent can be entirely waived if the research meets four conditions:^{14,46.116}

1. the research involves no more than minimal risk to the subjects;
2. the waiver or alteration will not adversely affect the rights and welfare of the subjects;
3. the research could not practicably be carried out without the waiver or alteration; and
4. whenever appropriate, the subjects will be provided with additional pertinent information after participation.

These criteria are often applied to pharmacoepidemiology research and other forms of research that rely on the use of pre-existing records. The controversial conditions here are whether the research risks are minimal and whether a waiver of informed consent will adversely affect the subjects’ rights and welfare. These are controversial because, in research that involves the use of medical records, the principal risk is the violation of the subjects’ confidentiality. A consensus about the proper application of these conditions requires a consensus about whether access to the patient’s medical record without the patient’s permission is a violation of confidentiality that is greater than minimal risk and violates the subject’s rights and welfare.

There are two competing answers to this question. The first relies upon a strict adherence to the principle of respect for autonomy. Accordingly, any unauthorized use of records violates confidentiality, presents more than minimal risk, and adversely affects subjects' rights and welfare. Hence, in all human subject research, the subject's informed consent could be perceived as an absolute requirement. Although this view follows from strict adherence to some research ethics codes,²⁴ this is not the view held by most contemporary researchers and ethicists.

Instead, a second interpretation allows for flexibility in the priority of the principle of respect for autonomy. Accordingly, some potential or even actual violations of confidentiality do not adversely affect the subject's rights and welfare or present more than minimal risk. This interpretation requires that we be able to determine to which kinds of information, if any, most people would be willing to grant access. For instance, at one extreme, research using information about patients' sexual activity or certain genetic characteristics might well be perceived as posing a greater than minimal risk. In such a study, obtaining that information without patients' consent might well have an adverse impact upon patients' rights and welfare, depending on the use of the information and the safeguards in place to protect access by third parties. In contrast, information about patients' age and blood pressure might seem to pose only minimal risks, even though blood pressure information could be more predictive of future disability status than the results of a genetic test. Obtaining information without patients' consent must be considered in 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. This first is to assure that the ethics board review is truly multi-disciplinary 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 PHARMACOEPIDEMOLOGY RESEARCH

There are several procedures available that can protect patient confidentiality. These methods allow patients to control who has access to information. At the time that clinical data are gathered, such as upon enrollment in a health system, patients can provide a "universal consent" to determine whether his or her medical record can be used for research. This term should not be construed to mean an informed consent to participate in research, because the patient is simply consenting to the generic use of his or her records and not whether to participate in actual protocols. A variation on this method is that patients can shield some aspects of their medical records from use in research. This is possible in some electronic record management systems. For example, patients could place into an electronic "black box" records of certain medications, such as antidepressants. Finally, at the time of the research, patients can be contacted to provide informed consent for the use of their archived records. However, there are two problems in applying these methods to pharmacoepidemiology 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, and therefore generalizability to the population that stands to benefit from the research.

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, such contact may be considered intrusive by the individual. Individuals may consider that their confidentiality has been violated if researchers must 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 ground, they all provide grounds for concern that a variety of procedures for protecting privacy may not be ideal.

Validity is a necessary precondition for all ethical research,²⁵ and research should not be done if it cannot answer the hypothesis it claims to test. In pharmacoepidemiologic studies that use archival records, methods that allow patients to control who has access to data can severely limit the validity of the research to be done. For instance, consider the procedure of universal consent, in which each patient is given the opportunity to remove his or her electronic medical records (such as Medicaid data) from use research. It is certain that at least some patients will opt out. The problem is that willingness to provide consent is generally not random, and varies in ways that may bias study results, as demonstrated in a study of consent bias in the Rochester Epidemiology Project.²⁶ In an ethics review board approved study at the Mayo Clinic, patients were mailed an educational brochure and a request for authorization to allow use of medical records for research. Characteristics of patients who did and did not provide written authorization were compared. Among persons returning the form, the refusal rate was low (3.2%), but the persons declining consent varied from the study population by age, gender, residence, and prior diagnoses, suggesting that the ability to opt out of databases creates a potential bias in the data.

Such practices may prohibit the evaluation of a key medication-adverse event association if the shielded information is in the pathway between the medication exposure and outcome of interest. For example, the results of a study of a drug-outcome association may be misleading if there is a large increased risk due only to an interaction between the study medication and the shielded antidepressant. The overall study results may show a low level association between the study drug and the

outcome. No interaction could be analyzed. Further, if all patients treated with antidepressants chose to withhold their medical records from any research, the drug-outcome study would show no association, since no data on patients experiencing the reaction would be included in the research data file.

When researchers attempt to contact all patients in a database to seek informed consent, some patients may be unavailable to provide consent because they have died, moved, or changed health plans. Those patients are likely to be distributed in a nonrandom fashion. The potential bias was demonstrated using hypothetical data from the Mayo Clinic Rochester Epidemiology Project.²⁷ Data available from all patients over a 50 year period showed a decrease in the population incidence of hip fracture. Data from only those patients known to be alive and able to give consent would produce results showing an increasing risk of hip fracture over time. This consent issue poses particular challenges in studies requiring long periods of exposure or followup, studies evaluating events of long latency, and the evaluation of intergenerational effects of medications.

The number of studies using archival records will likely increase with growing availability of electronic records and increasing interest in answering important drug safety questions. However, as the number of studies increases, there will undoubtedly be a decreasing consent rate if all studies require consent. Jacobsen showed a high consent rate among persons who returned consent forms, but only 79.3% of persons contacted returned the consent forms.²⁶ This return rate should be considered optimal, given the high credibility of the Mayo Clinic within Minnesota. Another study of consent, a drug safety study within a population of members of a Minnesota health plan, showed much poorer participation. In this study, with more representative results, only 19% of individuals contacted provided consent, and only 52% provided any response.²⁸

An additional problem is encountered in the conduct of large, multi-institutional case-control studies in which access to a large number of data must be reviewed in order to identify the cases and controls prior to contacting the appropriate

patients for consent. Ethics review boards typically waive the requirement for consent in the initial case finding review of records, and evaluate the consent used when patients are invited to participate in the study. Applying the current Common Rule framework to these studies requires separate review by ethics review boards from each participating institution of the same protocol. Issues raised by these ethics review boards and encountered in the review process may relate less to true local differences in the research environment than to the administrative differences of each institution's ethics review board process. Absent a more streamlined approach to the current ethics review board process, the time and cost of seeking multiple approvals discourage the conduct of these studies that may have important public health implications.

CURRENTLY AVAILABLE SOLUTIONS

These methodological challenges pose considerable obstacles to the conduct of pharmacoepidemiology research. For records based studies using data not directly identifying subjects, investigators have relied on the confidentiality policies governing the use of information in the individual institution. For studies using identifiable records, investigators receive guidance and direction, if they receive it at all, through a process of negotiation with local ethics review boards, whose task is to balance the requirements of the research design with the rights and welfare of prospective subjects. Because the tension between ethics requirements and the exigencies of pharmacoepidemiology research require this balancing process, in a very real sense the ethics of pharmacoepidemiology research is a negotiated agreement between investigators and one or more review boards. The available solutions to the methodological challenges outlined in the previous section, therefore, depend upon two factors. First, they depend upon the steps investigators can take in gathering and handling data. Second, they depend upon the degree to which review boards can and should be involved in research, and on their ability to review research in a manner that is both

competent and efficient. We examine each of these in turn.

The past several years have seen a rapid movement toward legislative protections for data privacy both in the US and internationally. These legislative approaches to protect the confidentiality of medical data provide potentially strong protections and safeguards on the creation and reuse of confidential information. For instance, the European Union (EU) directive that went into effect in October 1998 covers all information that is either directly identifiable or information from which identity could be inferred.²⁹ EU member states are currently bringing their laws into conformity with the directive, and tailoring these laws to the individual concerns and circumstances of their countries. The EU directive requires, first, consent for all uses of information beyond that for which the information was originally collected. Safeguards on the use and transfer of information are required as well. Each institution must have a data controller/data privacy officer, who is accountable for appropriate procedures and use of data within the institution. In addition, data cannot be transferred from a member state of the EU to another country outside the EU unless that country has safeguards at least as stringent as those of the EU. (This provision has led to revisions in law and/or regulation now under way in the US.) Notably, however, member states may grant deviations from some provisions of the directive for activities of substantial public interest. Interestingly, there is no mention of ethics review boards in the Directive. All research would presumably (i) be conducted with explicit consent, (ii) be conducted only with delinked records, or (iii) be exempted by a specific member state as a type of activity of substantial public interest.

For pharmacoepidemiology, a number of implications of the directive are of concern. For example, pharmacovigilance activities currently must be conducted using identifiable data. A requirement for patient consent would stifle the collection of a substantial proportion of cases and therefore hinder the ability to identify signals of drug safety problems. Furthermore, analysis of secondary information (from clinical trials or administrative databases) for research questions

not anticipated at the time patients signed consent, would not be possible without additional consent. Very little research could be conducted using secondary files from which direct patient identifiers have been deleted. This restriction is due to the broad definition in the directive of identifiable and “indirectly identifiable” data.

In the US, proposed legislation promises stricter scrutiny of research and tighter protections as well. Bills S.881, S.578, and S.573 all promise 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 most proposals, research using identifiable records would be covered and what constitutes “identifiable” information is, therefore, a critical factor determining the true scope of these laws.

One of the more novel proposals to emerge from these US bills has been the creation of strategies for addressing the ethical challenges created by the use of existing datasets. In these bills, several options would be available: review by an ethics review board, review by a committee established for review of protocols for archival records studies, or review and approval of a data confidentiality officer. Under this latter model, the data confidentiality officer would assure appropriate safeguards are in place, tailored to the local circumstances, and would also determine which studies could be approved, denied, or referred to an ethics review board. These last proposals may have considerable promise. Because oversight of electronic medical records confidentiality requires a more system-wide view than a study-specific review performed by review boards, a model that requires a data privacy officer has much to recommend it. That officer would be responsible for creation, maintenance, and compliance with procedures tailored to the needs and issues of the specific institution, and is responsible for maintaining records of all uses of identifiable data. The office would be organizationally isolated from functions in the organization with a vested interest in the use of the data. A data privacy officer could establish local policies that determine which studies might risk patient confidentiality and therefore require ethics review board review.

Because pharmacoepidemiology research often requires use of electronic records across large populations, studies may cross many different institutions (e.g., hospitals, physician practices) and multiple provider groups or insurers. In this kind of multi-institution research, a key that links the identifiable records with the code in the de-identified research files is maintained separately and cannot be accessed by the researcher.³⁰ The data management officer model provides one way to streamline the delinking process.

While these proposals hold some promise, there are also opportunities to improve the ethics board review process. Ethics review varies widely from country to country, and they may even be different within one country. In existing guidelines there is general agreement that protocol review by ethics review boards is valuable in principle. However, there is considerably less agreement about what kinds of pharmacoepidemiology research require this review.

For instance, the Royal College of Physicians position statement on ethics board review suggests that ethics board review is not necessary, even for linked studies, as long as investigators take appropriate precautions to safeguard the confidentiality of information.¹⁵ On the other hand, the Committee of the International Organization of Medical Societies recommends ethics board review for all epidemiology research, whether or not they involve identifiable data.¹⁶ In the middle are the recommendations of the International Society for Pharmacoepidemiology (ISPE)⁷ and the Common Rule^{14, 101, b4} which require ethics board review only when subjects are identifiable through linked data.

In some cases, it is not even the features of the research, but the source of funding that determines whether ethics board review is necessary. For example, as noted above, the Common Rule regulations apply only to research that is conducted using federal funds or research that is conducted in institutions that have agreed to follow these regulations voluntarily. The result is that while some researchers are required to apply for ethics review board approval, other researchers whose research present the same kinds of research risk are not. Although this distinction on the basis of funding source respects the limits of federal

authority in intrastate activities, it lacks moral force.

As examples of efficient protection of human subjects, the Common Rule and ISPE positions seem the most sensible position. This means that investigators and ethics review boards will at times need to negotiate the kinds of research that achieve standards such as "existing data" and minimal risk. However, this negotiation is a far better system to assure adequate subject protections for research than a system in which decisions are either entirely left in the hands of the investigators or made by others.

Nevertheless, this system of research oversight, and its heavy dependence on ethics board review, means that oversight can vary widely among institutions. This variability creates enormous administrative challenges for pharmacoepidemiology investigators, challenges that may be magnified in the case of multicenter research that crosses international borders. Certainly, sensitivity to local issues may be a desirable feature for the ethical review of research, particularly if institutions have special populations or circumstances that warrant special scrutiny of protocols. However, this variability may also be the result of variability in the quality of the ethics review board's skills and resources.

The ability of ethics review boards to review research in a manner that is both competent and efficient addresses issues of the training and certification of membership and resources for handling the volume of new and renewing research protocols. The Office of Extramural Research's 1998 evaluation of the ethics review board system suggested that among the chief causes of variability in the quality and efficiency of ethics board review are the limited resources available to ethics review boards and the variable competency of their membership. 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. However, these solutions do not adequately address a larger problem. Although ethics review boards may offer a reasonable procedural solution to ethics review, it is less clear how ethics review boards should make the sorts of decisions that are required of them. Specifically, it is not clear how ethics review boards and investigators should balance ethical and 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 idea of balancing is not new. Traditional approaches to balancing the ethical and methodological requirements of research typically use as their guide the research risks. In most guidelines, and the Common Rule is an excellent case in point, increasing risks to subjects requires increasing attention to full ethics board review and the informed consent process, including written documentation of informed consent.¹⁴ Seen in this light, there is a simple proportional relationship between research risks and subject protections such as informed consent. This relationship describes the degree of subject protections solely in terms of the balance of the risks and potential benefits to the subjects of the proposed research.

The problem, though, is that this relationship is too simple for the situation of pharmacoepidemiology 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. They are instead largely risks of another kind. The chief risk is the violation of confidentiality, which is really a civil, rather than a medical, risk.

We suggest that investigators and ethics review boards should consider an additional factor in this relationship: the value of the knowledge to be gained.^{14,46,111a} An ethical justification for this position begins, first, with the example of social services research. United States research regulations currently include an exception for studies designed to evaluate social programs.^{14,46,101} The implicit argument for this exception is that these social programs offer clear and evident value. They contribute in an important way to the social good. Studies designed to evaluate them, even if these studies bear all of the markings of “research,” are considered to be exempt from the requirements of ethics board review and subject informed consent that govern the ethical conduct of research. In a sense, the requirements of ethical research are suspended for studies that offer significant and generally agreed upon value.

This is an extreme case of balancing value against research risks. Indeed, it effectively removes research involving social programs from the purview of ethics oversight. This example is informative not only because it is so extreme, but also because studies of social programs have a great deal in common with pharmacoepidemiology 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 pharmacoepidemiology 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 project, 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 pharmacoepidemiology research because pharmacoepidemiologic re-

search, like any research, spans a continuum. Certainly studies of adverse drug reactions resemble closely the example of social program research. This is one standard, perhaps the highest standard, for a study’s potential to produce valuable knowledge. Phrased somewhat differently, the knowledge must be immediately relevant and applicable to the subjects who are being studied. In pharmacoepidemiology research, one example might be a study of adverse drug reactions among individuals taking a certain medication. Results of this research would have immediate consequences for the health of the patients, or “subjects,” for whom data is gathered.

Other studies may be done for private companies or organizations following vigorous methodological standards but where the findings would not be made public or shared with anyone outside the sponsoring organization. It is difficult to know how to balance concerns for privacy against the desire of private entities to obtain pharmacoepidemiology data. Studies like these should arguably be held to a different ethical standard because they do not hold the immediate possibility of clinically relevant knowledge that could be applicable to the people involved. The problem is that no public and national body exists to decide what kinds of research achieve this level of value.

The central ethical issue in pharmacoepidemiology research is deciding what kinds of project will generate generalizable knowledge that is widely available and highly valued and do this in a manner that protects individuals’ right to privacy and confidentiality. The problem is that these two ends differ in kind. The knowledge generated by pharmacoepidemiology is health related knowledge about such things as the risks and benefits of medicines. In contrast, individuals’ right to privacy is a matter of civil law. Although the two are frequently cast as in need of balancing, it may not be possible to weigh a certain amount of knowledge to be gained against a certain amount of confidentiality to be lost.

Instead, perhaps the most productive approach will be to determine what kinds of procedures and practices warrant crossing thresholds of confidentiality in the pursuit of valuable knowledge. Such a discussion should include research, but should not

by any means be limited to it. For example, society allows journalists wide access to gather and disseminate information, provided the journalist adheres to standards of practice (such as preserving the confidentiality of sources) and journalism is still viewed as a valuable instrument for preserving a democratic society.

Therefore, if the ethical requirement of informed consent is absolute and inviolable, then any balancing would be indefensible. However, this is not a tenable solution, nor is it a solution that would be consistent with the way that society responds to a need for valuable information in other settings. Further public discussion is needed to identify ways in which the policies and procedures for the protection of privacy and the maintenance of confidentiality are fair and consistent with the requirements imposed on other sectors of society.

REFERENCES

- Advisory Committee on Human Radiation Experiments. *Final Report*. Vol. 061 00000848-9. Washington, DC: Government Printing Office; 1995.
- Caplan AL. Twenty years after. The legacy of the Tuskegee Syphilis Study. When evil intrudes. *Hast Cent Rep* 1992; **22**: 29-32.
- Beecher HK. Ethics and clinical research. *New Engl J Med* 1966; **274**: 1354-60.
- Kahn JP, Mastroianni AC, Sugarman J. *Beyond Consent. Seeking Justice in Research*. New York: Oxford, 1998.
- Last JM. Guidelines on ethics for epidemiologists. *Int J Epidemiol* 1990; **16**: 226-9.
- Beauchamp T, Cook R, Fayerweather W *et al*. Ethical guidelines for epidemiologists. *J Clin Epidemiol* 1991; **44**: 151S-69S.
- Andrews E. Data Privacy, medical record confidentiality, and research in the interest of the public health. *Pharmacoepidemiol Drug Safety* 1999; **8**: 247-60.
- Organization for Economic Co-operation and Development. *Guidelines on the Protection of Privacy and Transborder Flow of Personal Data*, recommendation of the OECD Council, 23 September 1980.
- Mann RD, Bertelsmann A. The issue of data privacy and confidentiality in Europe—1998. *Pharmacoepidemiol Drug Safety* 1999; **8**: 261-4.
- Brody BA. *The Ethics of Biomedical Research: an International Perspective*. New York: Oxford University Press, 1998.
- Lynn J, Johnson J, Levine RJ. The ethical conduct of health services research: a case study of 55 institutions' applications to the SUPPORT project. *Clin Res* 1994; **42**: 3-10.
- 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.
- Levine RJ. *Ethics and Regulation of Clinical Research*, 2nd edn. Baltimore, MD: Urban and Schwarzenberg; 1986.
- Department of Health and Human Services. *Protection of Human Subjects*, Title 45 Part 46: Revised. Code of Federal Regulation; 18 June 1991.
- Royal College of Physicians. Research involving patients. Reprinted in: Brody H. *The Ethics of Biomedical Research. An International Perspective*. New York: Oxford University Press; 1990; 315-20.
- Council for International Organizations of Medical Sciences. International guidelines for ethical review of epidemiological studies. Reprinted in: Brody H. *The Ethics of Biomedical Research. An International Perspective*. New York: Oxford University Press; 1998; 225-32.
- Warren S, Brandeis L. The right to privacy. *Harvard Law Rev* 1890; **4**: 193-220.
- Fried C. Privacy: a rational context. *Yale Law J* 1968; **77**: 475-93.
- Avineri S, de-Shalit A. Individualism and Communitarianism. New York: Oxford University Press; 1992.
- Freedman B, Fuks A, Weijer C. In loco parentis. Minimal risk as the ethical threshold for research upon children. *Hast Cent Rep* 1993; **23**: 13-9.
- Grisso T, Appelbaum PS. Assessing Competence to Consent to Treatment. New York: Oxford University Press; 1998.
- Brett A, Grodin M. Ethical aspects of human experimentation in health services research. *J Am Med Assoc* 1991; **265**: 1854-7.
- Truog RD, Robinson W, Randolph A. Is informed consent always necessary for randomized, controlled trials? *New Engl J Med* 1999; **340**: 804-7.
- The Nuremberg Code; Reprinted in: Brody H. *The Ethics of Biomedical Research. An International Perspective*. New York: Oxford University Press; 1998; 213.
- Freedman B. Scientific value and validity as ethical requirements for research: a proposed explication. *IRB* 1987; **9**: 7-10.
- Jacobsen SJ, Zhiisen X, Champion ME *et al*. Potential effect of authorization bias on medical record research. *Mayo Clin Proc* 1999; **74**: 330-8.
- Melton LJ. The threat to medical-records research. *New Engl J Med* 1997; **337**: 1466-70.

28. McCarthy DB, Shatin D, Drinkard CR. Medical records and privacy: empirical effects of legislation. *Health Serv Res* 1999; **34**: 417–25.
29. 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.
30. General Accounting Office. *Medical Records Privacy: Access Needed for Health Research, but Oversight of Privacy Protections is Limited*, GAO/HEHS-99-55, February 1999.

27

National Medicinal Drug Policies: their Relationship to Pharmacoepidemiology

SUZANNE HILL

Discipline of Clinical Pharmacology, School of Population Health Sciences, Faculty of Medicine and Health Sciences,
The University of Newcastle, NSW, Australia

DAVID HENRY

Discipline of Clinical Pharmacology, School of Population Health Sciences, Faculty of Medicine and Health Sciences,
The University of Newcastle, NSW, Australia

INTRODUCTION

Pharmacoepidemiology exists primarily to quantify the effects of medicinal drugs in communities. Drug policies are the instruments used by governments (with varying degrees of success) to control the development, distribution, subsidization, pricing, and use of drugs in communities. Pharmacoepidemiology methods, such as pharmacoeconomics, drug utilization studies, adverse drug reaction monitoring, and formal analytical designs, are some of the tools that can be used to plan, monitor, and evaluate medicinal drug policies. It can be seen that drug policies are very relevant to pharmacoepidemiology, and vice versa.

National Medicinal Drug Policies have become a topical subject in recent years. A number of developing and developed countries now have comprehensive written policies that are implemented to varying degrees. It must also be recognized

that many countries have made significant progress in controlling the supply and use of modern drugs without recourse to written policies. However, national governments, district health authorities, and managed care organizations around the world have to deal with the rising costs of modern drugs, and the results of the suboptimal prescribing and consumption patterns that are common in so many communities. National policies that set standards to be followed by manufacturers, healthcare organizations, health professionals, and patients are very valuable, even in an era of "small government." Such policies have to operate in an environment shaped by budgetary pressures, community expectations, medical politics, corporate lobbying, and media coverage.

This chapter can do no more than scratch the surface of this complex subject; many specific issues are covered in more detail in other chapters (e.g., adverse reaction detection, drug utilization

studies, and pharmacoeconomics). Readers who wish to learn more are referred to the web page of the World Health Organization/Essential Drugs and Other Medicines (WHO/EDM) (<http://www.who.int/dap/edm.html>), which has an extensive range of technical reports, and to an excellent publication *Managing Drug Supply*.¹ We have drawn extensively on this work when compiling this chapter.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

The “problem” that needs to be addressed in relation to drug policies has been succinctly defined by Sven Hamrell² in a study of drug policy in South-East Asia.

The problem isn't so much in the drugs, which are medicines used to relieve sickness or cure disease. The problem lies in the way these drugs are dispensed, manufactured and marketed. The problem can be put simply: there is a disparity between the world's health needs and what drugs the world actually gets.

It is a sad indictment of our attempts to improve the health of the world's populations that, while the technologies and treatments needed to save lives and reduce suffering are technically feasible and available in many countries, they are denied to too many of those who need them most. The reasons are a lack of political will, an international desire to protect the commercial interests of some of the world's most profitable corporations, and significant failures of national drug policy development and implementation in many developing countries.

The problems that beset developing countries can be categorized under four broad and overlapping headings:

- poor quality and counterfeit drugs
- lack of availability of essential drugs for a population

- poor quality use of pharmaceuticals, e.g., overuse, misuse, polypharmacy
- unintended effects of international activities

It is entirely appropriate to give priority to the policy needs of developing countries. However, there are many examples of policy failures in developed countries as well, and we will highlight some of these in this chapter in addition.

POOR QUALITY AND COUNTERFEIT DRUGS

Drugs must be produced according to appropriate standards of good manufacturing practice. Although such standards have been developed and implemented in most developed countries, and are widely promulgated by the World Health Organization, many developing countries do not have the technical, financial, or human resources required for their implementation. Often criticized for their promotional activities in developing countries, trans-national pharmaceutical manufacturers have high quality manufacturing practices and can help maintain standards in developing countries. However, because many drugs are unaffordable (at international prices), there are major incentives for the development of local producers. Lack of technical expertise, and ineffective government inspection, may lead to substandard products being sold, and in some circumstances contaminated or counterfeit products becoming available. Hanif *et al.*³ and O'Brien *et al.*⁴ provide the two most recent descriptions of “epidemics” of renal failure in children due to diethylene glycol contamination of acetaminophen syrup (perhaps mistaken for the nontoxic propylene glycol). There are many other reports of diethylene glycol poisonings, dating back to the 1937 “epidemic” in the United States.⁵⁻⁹ In all cases, the contamination has been a result of the failure of manufacturing practice, or the failure of regulatory and surveillance systems to adequately control manufacturing practices.

Counterfeit drugs are another manifestation of problems of inadequate regulation and control of the quality of pharmaceuticals. A recent example was the release of a counterfeit meningococcal

vaccine during an epidemic of meningitis in Niger.^{10,11} As described by Pecoul *et al.*,¹² the problem was first noted when a team from Mediciens Sans Frontieres working with local health workers noticed that vaccines from Nigeria, labeled as a product from Pasteur Merieux, had an unusual appearance. Subsequent investigations identified them as counterfeit. It was estimated that approximately 60 000 persons were vaccinated with the false vaccine; the scale of the production would have required a large scale manufacturing facility. In other words, this was large scale organized crime, certainly not an isolated example.

There seem to be two general groups of counterfeit products—organized illegal circuits that manufacture copies of known trademarked products, and nonorganized production of small quantities of inadequate or substandard products, including generic drugs. Shakoor *et al.*¹³ described a study of chloroquine products available in Thailand and Nigeria which found that 36.5% of the products sampled were substandard. In general, this appeared to be due to substandard manufacturing processes and problems with decomposition of the products, rather than fraudulent manufacturing. Nazerali and Hogerzeil¹⁴ described a similar problem in an investigation of the quality of essential drugs in Zimbabwe; they identified initial quality problems with ampicillin, retinol, and ergometrine, as well as problems with the stability of ergometrine injections over time.

Ensuring that pharmaceutical products are of satisfactory quality not only requires adequate controls on manufacturing. An adequate distribution system for drugs is critical to ensure that drugs reach the user in good condition. Degradation of products during distribution has been most extensively documented for vaccines; this has been a problem in both developed and developing countries. Standards for vaccine handling have been developed by WHO, initially to allow for the transport of the most fragile of the essential vaccines, oral polio vaccine.¹⁵ Studies in South Africa,¹⁶ Italy,¹⁷ Australia,^{18–21} Malaysia,²² India,²³ and Nigeria²⁴ document the difficulties of maintaining an adequate cold chain in different environments. The problems that were identified

ranged from the central storage facility having inadequate monitoring systems to refrigerators in pharmacies and doctors' surgeries being inadequate for storage of vaccines immediately prior to their administration.

Serious problems can also occur during bulk transport. Hogerzeil *et al.*²⁵ examined drugs that were shipped by sea or land from a donor agency to recipient countries. They found that the drugs in the shipment were exposed to much higher temperatures and humidity than was recommended by their manufacturers. However, this only affected the clinical effectiveness of two of the products.

LACK OF AVAILABILITY OF ESSENTIAL DRUGS

The chronic mismatch between drugs that are available in developing countries, and the diseases and populations requiring treatment, led to the development of the "essential drugs" concept. The history of this important movement has been described in detail by Kanji *et al.*²⁶ "Essential drugs," as defined by WHO in 1975, are "those considered to be of utmost importance and hence basic, indispensable and necessary for the health needs of the population. They should be available at all times, in the proper dosage forms, to all segments of society." The first model list of essential drugs was prepared by WHO in 1977; it is now in its tenth edition. Despite the policy having been adopted in many countries, Pecoul *et al.*¹² consider that access to essential drugs of adequate quality appears to be getting worse, not better, in many settings. This is illustrated when one considers the consumption of pharmaceuticals in various countries. Developed countries account for approximately 16% of the world's population, and, in 1990, accounted for approximately 72% of consumption, with the US and EU being the largest markets. Developing countries account for approximately 77% of the world's population, and in 1990, consumed 19% of the pharmaceuticals. This has declined from 24% in 1975.²⁷

There have been many studies that describe the lack of availability of essential drugs in developing and developed countries. In Rwanda, for example,

Habiyambere and Wertheimer²⁸ estimated that 70% of the population did not have access to essential drugs. Another manifestation of this problem can be excessive and unregulated supply of the wrong drugs, as described by Sri-Nyernyaung²⁹ in Thailand. The reasons for lack of essential drugs may vary from setting to setting, ranging from lack of production because the drug is unprofitable (such as meglumine antimoniate for leishmaniasis¹²), to because the drugs are too expensive, such as the widely discussed antiretrovirals for the treatment of HIV. Other more complex reasons have been described by Foster³⁰ and may relate to the structure of the local pharmaceutical market and the various components of the supply and distribution chain.

The mismatch between what is available and what is needed can lead to a range of strategies to obtain essential drugs. Kandela³¹ illustrates this with the example of the effect of the economic embargo on Iraq. As the supply there of essential drugs became compromised, residents with relatives and friends overseas ensure that any visitors' suitcase contains supplies of basic drugs such as antibiotics and thyroxine. Kent and Glatzer³² noted the chaos that developed in Bosnia, with the media coverage determining the supply of particular medicines. If the story of the week emphasized a shortage of anesthetic agents, the hospital concerned would be overwhelmed with inappropriate donation of expensive anesthetic agents at the expense of more essential products. (Drug donations are discussed in more detail below.)

Lack of access to essential drugs is not only a problem in developing countries. In developed countries, vulnerable populations can be without access to essential drugs, usually because of avoidable failures of government policy. The most spectacular examples of this are in the US, where a significant proportion of the population lacks guaranteed access to affordable supplies of the drugs that they need.³³ Attempts to contain drug costs at a state level or in managed care organizations can have unintended effects. In an earlier study, Soumerai *et al.*³⁴ described the impact of a three-month prescription payment limit or "cap" on the use of psychotropic drugs and acute mental

health care by patients with schizophrenia. Not surprisingly, the "cap" resulted in an immediate reduction in the use of psychotropic medication by the population. With decreased consumption of essential medications, the use of acute mental health services (and the associated costs) increased. Rabon *et al.*³⁵ describe another manifestation of this type of problem in their survey of the availability of cancer medication in pharmacies in South Carolina. They found that only 25% of a list of cancer medications were available to be dispensed from 90% of the pharmacies surveyed.

In many cases, the lack of availability of essential drugs is due to the high prices being requested by manufacturers. This mismatch of demand and affordable supply is an important example of market failure, and the inadequacy of government intervention in many countries. International comparisons of prices of pharmaceuticals are difficult and are influenced by fluctuating exchange rates, variable purchasing power of currencies, and wages. However, Health Action International, a consumer organization that aims to promote the rational use of drugs, published price comparisons of 12 commonly used drugs in five Asian countries and Canada (*HAI News*, December 1995). Enormous price variation was documented. For example, the price of 100 × 150 mg ranitidine tablets ranged from \$US3 in India to \$US150 in Indonesia, presumably due to different manufacturers and suppliers negotiating different prices, plus the variability in retail systems.

Finally, it is important to recognize that "essential" drugs only deserve that title so long as they are used in an appropriate manner. There are many examples of the misuse of drugs on the WHO's model list, and ensuring appropriateness of use is just as important as establishing stable systems for selection, procurement, and distribution.

POOR QUALITY USE OF PHARMACEUTICALS

The third aspect of the policy problem that needs to be considered is the poor quality *use* of pharmaceuticals. This can be divided into three

areas: overuse of pharmaceuticals, underuse, and misuse. The problems of overuse have probably received most attention and there is an extensive literature in this field. The overuse and inappropriate use of antibiotics is a problem internationally.³⁶ Most surveys in developing countries show that antibiotics are prescribed in 35–60% of clinical visits, although they are appropriate in less than 20%.³⁷ The size of the antibiotic market in developing countries has been estimated to be double that of developed countries.³⁸ The factors that contribute to this include free availability of products combined with poor prescribing and information. The resulting significant problems of antibiotic resistance have been extensively documented. Ironically, many patients are unable to afford more than 1 to 2 days' treatment. The perverse situation arises in which those who do not need the drugs receive a brief course, which may encourage antibiotic resistance, while those who really need antibiotics receive a totally inadequate course of treatment.

Overuse of drugs in developing countries is not restricted to antibiotics. Nonsteroidal anti-inflammatory drugs have been described as being inappropriately used in Brazil, where expenditure on one agent (diclofenac) exceeds that on almost any other drug—a clear example of the misalignment of community needs and clinical practice.³⁹ Elzubier and Al-Shehry⁴⁰ have documented the excessive use of vitamins in diabetic patients in Saudi Arabia.

Medication practices vary not only by country, but by area within a country, so that descriptive studies of local prescribing patterns are essential to determine the precise nature of the problem in any given locality.

Hogerzeil *et al.*⁴¹ have described a standard set of indicators that were developed for the measurement of drug use in developing countries, but these can also be applied to a developed country setting, and these are summarized in Table 27.1.

In developed countries, there are other manifestations of the problems of overuse and polypharmacy. Use of greater numbers of drug therapies has been found to be associated with an increased risk of developing adverse drug reactions.⁴² Polypharmacy in the elderly has been shown to be

Table 27.1. Core drug-use indicators

<i>Prescribing indicators</i>	
Average number of drugs per encounter	
Percentage of drugs prescribed by generic name	
Percentage of encounters with an antibiotic prescribed	
Percentage of encounters with an injection prescribed	
Percentage of drugs prescribed from essential drugs list or formulary	
<i>Patient care indicators</i>	
Average consultation time	
Average dispensing time *	
Percentage of drugs actually dispensed	
Percentage of drugs adequately labeled	
Patients' knowledge of correct dosage	
<i>Health facility indicators</i>	
Availability of a copy of essential drugs list or formulary	
Availability of key drugs	

* Time that personnel dispensing drugs spend with patients.
 Source: Hogerzeil *et al.*⁴¹

associated with urinary incontinence, delirium,⁴³ and syncope.⁴⁴ Studies in a variety of settings have shown that patients over 65 years of age use an average of two to six prescribed medications and one to 3.4 nonprescribed medications.⁴⁵ In nursing homes in particular, overuse of psychotropic drugs has been shown to be a problem.⁴⁶

Equally problematic, however, is the underuse of appropriate medications in this population. This is probably best documented for the lack of use of β -blockers following myocardial infarction,⁴⁷ although it has also been suggested to be the case for the use of diuretics in hypertension,⁴⁸ and (more controversially) the use of hormone replacement therapy.⁴⁹

UNINTENDED EFFECTS OF INTERNATIONAL ACTIVITIES

Although this chapter is primarily concerned with national drug policies, we cannot ignore the impact of some important global trends. One that we wish to highlight here is the difficulty created for national governments, and aid workers, by unsolicited donations of drugs during national emergencies caused by natural disasters or acts of war. Although these donations are nearly always

well intended, they can create many problems, as has been well documented.⁵⁰ Hogerzeil *et al.*⁵⁰ give a number of examples, including seven truckloads of expired aspirin tablets delivered to Eritrea that took six months to burn, and the arrival of contact lens solution, appetite stimulants, and lipid-lowering drugs in a war zone in the Sudan. Sometimes donated drugs have been close to expiry and in poor condition, labeled in foreign languages, and quite inappropriate to the recipient country's short to medium term needs. The work involved in disposing of unwanted pharmaceuticals can tie up key personnel for long periods, and may be very costly if it involves either burying the pharmaceuticals in concrete filled drums or high temperature incineration.

The motives behind drug donations may be altruistic (although perhaps not always—see below), but good intentions do not invariably translate into good deeds. One perverse incentive is the tax relief in the US that can be earned on the commercial value of drug donations. This has led to the cynical view that some manufacturers can make more money by donating expiring stock than trying to sell it to their normal clients. One difficulty for recipients of donations is that they do not want to appear ungrateful as it may compromise future, more appropriate, aid, and so vociferous complaints are uncommon.

The World Health Organization, in collaboration with a number of agencies, has produced detailed guidelines for drug donations (WHO/DAP 1996), and for the safe disposal of unwanted pharmaceuticals (WHO/EDM 1999). The guidelines make it clear that the responsibility of donors is to ensure that the recipients have been allowed to specify their needs, that the donations are appropriate, in good condition, have a long shelf-life, and are labeled in a way that enables them to be used effectively and safely in the field. If in doubt, cash donations will always be acceptable. It is likely that countries with well-established national drug policies will be well positioned to make the best use of appropriate drug donations.

The second major international influence comes in the form of the agreements on Trade-Related Intellectual Property Rights (TRIPS) mediated by the World Trade Organization (WTO). Through

these agreements, member states are expected to guarantee intellectual property rights for a period of 20 years.⁵¹ These agreements have the potential to limit the access of developing countries to new health technologies at affordable prices, as they might be denied the right to manufacture drugs locally at prices much lower than will be demanded by transnational corporations. This issue received a great deal of publicity during 1999 because of attempts by the US pharmaceutical industry to block the local production of cheap antiretroviral drugs in South Africa, where the prevalence of AIDS is very high. Much of the argument has centered on the legal interpretation of the WTO treaty obligations. However, most observers consider that TRIPS does not actually remove the right of countries to force compulsory licensing, i.e., forcing an international manufacturer to give a license to a local manufacturer to produce a drug that is essential to the national interest. This position was admitted by the US Trade Secretary in late July 1999, after an effective campaign by AIDS activists in the US that was aimed at Vice-President Al Gore. A coalition of organizations including the US-based Consumer Project on Technology, *Médecins Sans Frontières*, and Health Action International, held a meeting in early 1999 to explore ways of introducing compulsory licensing more widely as a means of improving the access of developing countries to the drugs that they need. Although this issue may be seen as one grounded in obscure interpretations of international law, it is clearly an important argument and one that will have major implications for national drug policies in the future.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

DEFINING NATIONAL MEDICINAL DRUG POLICIES

Public “policies” can range from having written documents that express intent on particular issues, to a complex process in which values, interests, and

resources compete through institutions to influence government actions. They should be distinguished from “politics,” although increasingly politics influence policy, as the costs of drugs continue to rise and restrictive policies are seen to limit access to drugs and to reduce corporate profits.

Health policies should reflect, and ideally meet, health needs. As pharmaceuticals are recognized to play a major role in meeting health needs, medicinal drug policies need to be part of health policies and use of drugs in a community should be consonant with overall health goals. This requires “careful analysis of existing and desirable relationships between health needs, drug demands and drug sales, within fluctuating political, economic and social pressures. These are often in conflict with each other because of the different and sometimes competing interests of the groups involved.”⁵²

Since the mid-1970s, national medicinal drug policies (NMDPs) have been developed in many countries to ensure access to pharmaceutical products of adequate quality, safety, efficacy, and cost-effectiveness. These developments have received strong encouragement, and great practical support, from the World Health Organization, through its Action Program on Essential Drugs (recently restructured as the Program for Essential Drugs and other Medicines). As a result, drug policy development has become recognized as a legitimate topic for study, not just the preserve of governments.

Ideally, NMDPs should provide the framework to coordinate the activities of the pharmaceutical sector, the public and private sectors, NGOs, donors, and other stakeholders. They have been described as a “guide for action.”¹ It is not only developing countries that should pursue such an approach; many developed countries have similar aims, but few have produced formal documents. Increasingly, experts are advocating that a policy document, even where the elements of a comprehensive NMDP already exist, can be of value, especially as part of any evaluation process.

OBJECTIVES OF NATIONAL MEDICINAL DRUG POLICIES

The objectives of an NMDP will vary from setting to setting, but in general terms, are likely to

include the following:¹

- to make essential drugs available and affordable to those who need them
- to ensure the safety, efficacy, and quality of all medicines provided to the public
- to improve prescribing and dispensing practices and to promote the correct use of medicines by health care workers and the public
- to ensure an appropriate balance of local production and importation of pharmaceutical products
- to build and maintain human capacity to ensure the sustainability of the NMDP.

Clearly, the extent to which these components are included in an individual country’s NMDP will vary depending on the country’s particular needs, and there are many examples of this variation. This is illustrated by Table 27.2, which summarizes the core objectives of national medicinal drug policies in a developing and a developed country (using Australia as an example).

It is immediately apparent that there are similarities and differences between these two sets of goals, and both are revealing. Assuring the efficacy, safety, and quality of medicines is a

Table 27.2. Comparison of policy objectives in developing and developed countries

Developing Country (from reference 1)	Developed Country (Australia)
● to make essential drugs available and affordable to those who need them	● to achieve affordable access to a satisfactory range of cost-effective drugs
● to ensure the safety, efficacy, and quality of all medicines provided to the public	● to ensure the safety, efficacy, and quality of medicines provided to the public
● to improve prescribing and dispensing practices and promote the correct use of medicines by health workers and the public	● to achieve quality use of medicines
	● to encourage the development of a successful pharmaceutical industry

common goal, but is much easier to achieve in developed countries because of the high quality of production facilities that usually exist in these countries. Both quality assurance procedures (maintenance of good manufacturing practices, close examination of data packages) and testing of samples can be carried out in a rigorous manner. As noted earlier, failure of these systems in developing countries may lead to drug contamination, substandard drugs, or counterfeit drugs. However these processes are labor intensive, and expensive, and a large investment may not be a priority when the health budget is insufficient to meet the most basic needs of the community.

Relevant to this consideration is the objective of any industry policy that is in place. This is explicit in the Australian policy. Most governments in developed countries appear to give a high priority to investment by large pharmaceutical manufacturers (although sometimes the reasons for this lack clarity), and also encourage the local production of generics. In some cases, the large international companies own the generic manufacturers. The situation in developing countries is different. It is often unclear whether a government should encourage local generic manufacture (risking quality failures), investment in manufacturing capacity by international companies (which may provide employment but do little to meet the health needs of the local community), or the development of sound procurement practices using the international market for essential drugs (most suppliers are based in Europe).

The affordability and accessibility of important drugs is an essential issue in any drug policy. Most developed countries (the US is a notable exception) have universal insurance programs that guarantee access to heavily subsidized drugs. Generally, the distribution and sales of drugs within such programs are handled fairly efficiently by the private sector.

Insurance programs are less common in developing countries, however, largely because of their great expense, and distribution systems that are taken for granted in developed countries often fail. A variety of procurement and distribution systems has been employed, with technical details that are beyond this scope of this chapter (see reference¹

for details). Finding the right public/private sector mix is a challenge; poor quality manufacture, theft, improper storage, and expiration due to poor stock control can lead to losses of 40% or more of the total value of purchases.¹

Achieving quality use of medicines (a less pejorative term than “rational drug use”) must be a priority in both developing and developed countries. In the former, the targets for educational campaigns and behavior change strategies include a wide range of health workers (including unlicensed drug sellers) and the public. In developed countries, most programs are aimed at prescribing doctors, and a few at consumers. However, essentially the aims are the same: to reduce the adverse effects associated with excessive drug use, to ensure that adequate treatment is given to those that need it, and to contain expenditures. In many ways, developing countries are more advanced in investigating and implementing techniques for improving drug use practices.⁵³ General methods for modifying physician prescribing are discussed in Chapter 30, and Chapter 31 discusses the use of drug utilization review as a specific and common technique used toward that end. In the remainder of this chapter, we will discuss the use of National Medicinal Drug Policies for this and other purposes.

CURRENTLY AVAILABLE SOLUTIONS

THE IMPLEMENTATION OF NATIONAL MEDICINAL DRUG POLICIES (NMDPs)

The components that are necessary to implement NMDPs have been defined by WHO:⁵⁴

- an appropriate legislative and regulatory framework
- a system for choosing appropriate drugs
- a mechanism for ensuring supply and distribution
- a program for improving rational use of drugs
- appropriate economic strategies for the pharmaceutical sector
- the development of human resources

- a system for monitoring and evaluating the impact of the policy
- research—drug research and development, and operational R&D.

As mentioned in the section on objectives of NMDPs, the correct policy mix will vary according to the stage of development and economic state of a country. However, experience indicates that successful implementation requires a strong central group with the necessary expertise, concentration on some key issues (it is not possible to achieve everything), public endorsement by senior politicians (the president/prime minister or health minister should “launch” the policy), international support, and a thick skin (the industry and local medical association may object on the grounds that the policy will reduce profits and clinical freedom).

A number of the features of NMDPs have been covered in previous sections of this chapter and there is insufficient space to do justice to all of the topics listed above. However, we would like to cover some selected implementation issues, including the growing interest in the use of formal economic analyses in the selection and pricing of drugs within a NMDP.

Legislative And Regulatory Frameworks

A key step for national drug policies is the drafting and implementation of an appropriate law that regulates the supply of pharmaceuticals. In developed countries, drug laws have usually evolved over many years. For example, in the US, the first national drug law was passed in 1906 and this was then modified following public health “disasters” such as the contamination of sulfonilamide in 1938 and thalidomide in the early 1960s (see also Chapter 1). The aims of most drug laws are similar: to regulate the supply of pharmaceuticals (including import and export), and to ensure that the forms in which pharmaceuticals are supplied meet accepted standards. Most drug laws do not try to regulate the way in which pharmaceuticals are used, by either the prescriber or the consumer.

In developing countries, drug laws have often been developed in the recent past, based on the

WHO Model Drugs Law, for example in Laos and some of the eastern European countries. To be effective, such legislation has to be supported by appropriate enforcement strategies. This may appear to be self-evident, but there are many examples where legislation has been implemented without enforcement, particularly relating to provisions about who can supply pharmaceuticals to the consumer. In Thailand and Vietnam, for example, despite the existence of national drug laws, there is as yet no effective control over the supply of antibiotics to the consumer, which contributes to major problems of inappropriate use.

It is not altogether clear what makes drug regulation effective. The WHO estimates that, of its member states, only 1/6 have an “effective” drug regulatory system.⁵⁵ It is possible to identify some factors that will contribute to ineffective drug regulation, such as lack of appropriate technical expertise, and inefficient or corrupt administration. Solutions to these problems are more difficult.

Systems For Choosing Appropriate Drugs And Ensuring Access

The Essential Drugs Program has probably been the most effective solution that has been implemented for choosing appropriate drugs. The advantages of using a limited list are summarized in Table 27.3. For example, Hogerzeil *et al.*⁵⁶ documented the impact of the introduction of an Essential Drugs Program in Yemen, that improved availability, use, and knowledge in relation to a selection of essential medicines. One of the difficulties, however, has been that in many settings, industry has not accepted the Essential Drugs Program and it has not been prepared to produce essential drugs at affordable prices.⁵⁷

At the International Conference on National Medicinal Drug Policies held in Manly, Australia, in 1995, the policy options for improving access to essential drugs were discussed. The options identified included price control, price competition, price awareness, and health insurance.⁵⁸ Although a number of examples appeared to be successful, the conditions for success of any particular strategy are still not fully understood.

Table 27.3. Advantages of adopting a limited list of essential drugs¹

Supply

- easier procurement, storage, and distribution
- lower stocks
- easier dispensing.

Prescribing

- training more focused and therefore easier
- more experience with fewer drugs
- no irrational treatment alternatives
- focused drug information
- better recognition of adverse drug reactions.

Cost

- lower prices, more competition.

Patient use

- focused education efforts
 - reduced confusion and increased adherence to treatment
 - improved drug availability.
-

Systems for choosing between drugs are developing. The “VEN classification” proposed by the WHO⁵⁹ represents a logical approach, as it provides a basis for choosing between a range of medications, of different types, and for different indications, when working under budgetary constraints. Briefly, drugs are classified as vital, essential, or nonessential, and this establishes a hierarchy for managing choices. When drugs are genuinely different from each other, those that are classified as “vital” should be selected ahead of those that are “essential.” When drugs are similar in their actions and indications (i.e., in the same VEN category), the decision should be based on an assessment of their comparative cost effectiveness (see Chapter 35). When the drugs are identical, the only decision criterion is the acquisition cost. As will be seen in the section on economic considerations, this approach, advocated by the WHO for use in developing countries, is very similar to the selection processes used in the operation of the Australian Pharmaceutical Benefits Scheme.

A Program For Achieving Quality Use Of Medicines

A key component of any NMDP is a program to encourage the quality use of medicines. Factors

that influence use of pharmaceuticals are complex and may be culture specific. A number of strategies for changing professional practice have been studied, for example the provision of standard treatment guidelines, the use of educational seminars, and more intensive approaches such as “academic detailing.” The most recent reviews of the various approaches suggest that a combination of methods is usually required to have sustained impact on practice.^{60,61} (See also Chapters 30 and 31.) The methods for changing consumer practices are still being defined.

The provision of adequate information about the efficacy and safety of pharmaceuticals is regarded as an important component of any program that is aimed at improving use of pharmaceuticals. Premarketing evaluation of drugs can help provide information about the efficacy of products, but it is well recognized that premarketing data do not provide the complete picture of the safety of products (see also Chapter 1). Surveillance systems for adverse drug reactions can be an important contributor to the information base as well as to other aspects of national medicinal drug policies. For example, spontaneous reporting systems can contribute to the education of prescribers and hopefully improve the quality of use of pharmaceuticals (see Chapters 10 and 11). In Australia, the hepatic side-effects of flucloxacillin that were identified through the spontaneous reporting system were publicized in the quarterly adverse drug reactions bulletin. Formal investigation of the problem (using case-control methodology) identified the dose and duration of therapy and the age of the patients as major risk factors for developing the reaction.⁶² This allowed appropriate advice to be issued regarding the usage of the product, and eventually there was a marked decline in usage.

Appropriate Economic Strategies for the Pharmaceutical Sector—the Example of Australia

The use of formal economic analysis (see Chapter 35) in the selection and pricing of drugs has been a controversial issue for a number of years. However, recently there has been a surge of interest; a

number of countries are considering the implementation of this approach and have published guidelines to assist the pharmaceutical industry in making applications for listing on their national formularies of subsidized products.

Most experience of this approach to formulary management has been gained during the operation of the Australian Pharmaceutical Benefits Scheme (PBS). The PBS and the processes used for evaluating drugs have been described in a number of articles.^{63,64} The use of economic analysis in the choice and pricing of pharmaceuticals is a central activity of the Australian national drug policy.

The PBS is a comprehensive, publicly funded, insurance program that reimburses pharmacists for the costs of a selected range of prescription drugs. Drugs are placed in different categories of access (restricted benefit, authority required), based on evidence of their comparative efficacy and cost-effectiveness in defined patient groups. The decisions to place new drugs in the PBS are made by the Commonwealth Health Minister on the advice of the Pharmaceutical Benefits Advisory Committee (PBAC). The PBAC receives advice from an economics subcommittee regarding the validity of the economic analyses contained in submissions from the pharmaceutical industry. The PBAC has a strong preference for using data from randomized controlled clinical trials as the basis of its judgement about the comparative efficacy of drugs. The guidelines for economic analysis released by the Department of Health and Aged Care (DHAC)⁶⁵ encourage the sponsors of new drugs to conduct a preliminary economic analysis, based on the results of randomized trials, before conducting a modeled analysis in which assumptions are made about long-term health effects and costs when the drugs are used in Australia. The database of submissions to the PBS that is maintained by the DHAC contains details of more than 300 submissions from the industry. Eighty-six percent of submissions were based on the results of randomized clinical trials, of which one quarter included the results of meta-analyses⁶⁶ (see Chapter 38). This is the most extensive experience in the world of the use of trial data in the assessment of the comparative efficacy and cost effectiveness of any healthcare intervention.

The key decisions that have to be made when a drug is considered for listing are whether it is superior to its comparators in clinical terms (improved efficacy, greater safety, both, or neither). The net costs associated with its use are then related to the assessment of comparative clinical performance. If there is no worthwhile difference in efficacy or safety, the new drug will receive the same price as the comparator (an established drug for the same indication). If the new drug is superior, the additional clinical benefits have to be related to the net costs in an incremental cost-effectiveness ratio. Whether this ratio represents “value for money” is the judgement of the advisory committee, and the decisions are sometimes quite controversial. It should be noted that factors other than cost-effectiveness are considered in the decision-making process, such as clinical need, equity of access, “rule of rescue,” and total cost to the health-care system. The complexity of this process reflects the true difficulties of making this sort of decision, and the provision of evidence at different stages helps to make the issues more clearcut.

The overall impact of this policy is difficult to judge, but one effect on drug pricing is fairly clear. The use of “cost minimization” to eliminate cost differences between very similar drugs has considerable social value, as it reduces expenditure that has no prospect of bringing greater health benefits. It also minimizes the need for educational efforts directed at doctors that are intended to shift prescribing between very similar drugs for purely economic reasons. An example of the effect of this policy is given in Figure 27.1, where the prices of commonly used (but very similar) nonsteroidal anti-inflammatory drugs in Australia and the United Kingdom are compared. The difference in average price is not the issue, as this is influenced by the exchange rate at the time the study was carried out. However, the notable difference between the two countries is in the price variation across this class of very similar compounds. Undoubtedly, there are benefits to the community in pursuing a policy that minimizes price differences across classes of equivalent drugs. This will apply in both developing and developed countries.

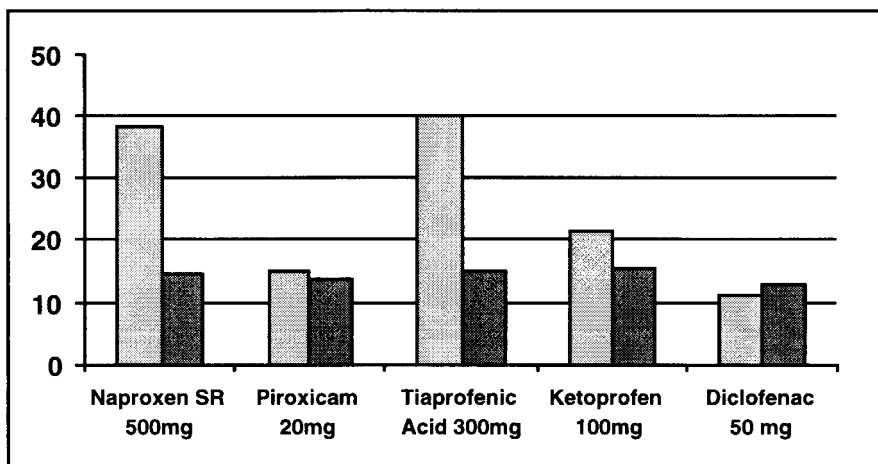


Figure 27.1. The impact of cost minimization on drug prices. A comparison of the prices of commonly used and similarly performing NSAIDs in Australia (dark shading) and the United Kingdom. The comparisons are based on the prices listed in the Australian Schedule and British National Formulary in the last quarter of 1997. The Australian prices were adjusted to the maximum quantities provided by a UK prescription. Prices are in \$AU, using an exchange rate of \$1AU = £0.40.

THE FUTURE

As countries become increasingly concerned with costs of drugs and with their proper use, it is likely that there will be increasing attention paid to the development of National Medicinal Drug Policies. The development of these policies will modify the types of research that pharmacoepidemiologists will be able to perform.

In addition, as these programs are expanded and developed, it is important that their effects are rigorously evaluated. There are several obvious roles for pharmacoepidemiologists in this process. These include drug utilization studies as part of a program to improve the quality of use of medicines (see Chapter 29), designing interventions to improve prescribing (see Chapters 30 and 31), the conduct and evaluation of pharmaco-economic studies (see Chapter 35), and identifying and quantifying adverse drug reactions (see Chapters 10–25). In developing countries, there is still a need for adverse reaction reporting as part of process to identify failure in the quality of production. A range of indicators of quality drug use have been suggested by the WHO, and by some countries. However, as with any system of assessment, there are virtues in simplicity.

Table 27.4 lists a few basic indicators that can be used to assess the degree of implementation of national drug policies. It should be stressed (again) that their use should not be confined to developing countries—they have relevance across all stages of economic development and there are few countries that will perform well on all of these indicators.

Pharmacoepidemiologists also have the potential to make major contributions to policies, by recognizing policy issues in the environments in which they work, measuring the impact of National Medicinal Drug Policies, and contributing to policy development. Many pharmacoepidemiologists already have considerable influence with governments and pharmaceutical manufacturers. That influence can be used to persuade governments to develop better policies, and to commit resources to policy research. It could also be used to persuade governments and manufacturers to respect the rights of developing countries and to recognize the legitimacy of their efforts to improve the access of their populations to essential drugs. Through these initiatives, both in research and policy, pharmacoepidemiologists can play an important role in improving the risk/benefit balance of drug use in their communities.

Table 27.4. Indicators used in the world drug situation to assess the effectiveness of national drug policies (modified from WHO 1988)

Indicators of availability

- The existence and use of an essential drug list (or selective national formulary)
- the extent of an operating system for procurement
- the extent of an operating system for distribution
- the extent of quality assurance
- the extent of regulatory mechanisms
- the extent of coverage.

Indicators of rational use of drugs

- The existence of a functioning and independent system that provides objective information on drugs to health workers and patients
- the existence of a system of continuing education for all types of personnel dealing with drugs
- the existence of a monitoring system for adverse drug reactions.

Indicators of the commitment of government

- The existence of a national medicinal drug policy.
-

REFERENCES

1. Quick JD, Rankin JR, Laing RO, *et al.*, eds. *Managing Drug Supply: the Selection, Procurement, Distribution and Use of Pharmaceuticals*, 2nd edn. West Harford, CT: Kumarian Press, 1997.
2. Dag Hammarskjöld Foundation. *Prescription for Change. National Drug Policies, Social Transformation and the Media*. Philippine Centre for Investigative Journalism, 1992.
3. Hanif M, Mobarak MR, Ronan A, *et al.* Fatal renal failure caused by diethylene glycol in paracetamol elixir: the Bangladesh epidemic. *Br Med J* 1995; **311**: 950–1.
4. O'Brien KL, Selanikio JD Hecdivert C, *et al.* Epidemic of pediatric deaths from acute renal failure caused by diethylene glycol poisoning. *J Am Med Assoc* 1998; **279**: 1175–80.
5. Geiling EMK, Cannon PR. Pathological effects of elixir of sulfonilamide (diethylene glycol) poisoning. *J Am Med Assoc* 1938; **111**: 919–26.
6. Bowie MD, McKenzie D. Diethylene glycol poisoning in children. *S Afr Med J* 1972; **46**: 931–4.
7. Cantarell MC, Fort J, Camps J, Sans M, Piera L. Acute intoxication due to topical application of diethylene glycol. *Ann Intern Med* 1987; **106**: 478–9.
8. Pandya SK. An unmitigated tragedy. *Br Med J* 1988; **297**: 117–9.
9. Okuonghae HO, Ighogboja IS, Lawson JO, Nwana EJC. Diethylene glycol poisoning in Nigerian children. *Ann Trop Paediatr* 1992; **12**: 235–8.
10. Pinel J, Varaine F, Fermon F, Marchant G, Marioux G. Des faux vaccins anti-meningocoque lors d'une epidemie de meningite au Niger. *Med Maladies Infect* 1997; **27**: 1–563.
11. World Health Organisation. Fake drugs: a scourge on the system. *WHO Drug Inf* 1995; **9**: 127–9.
12. Pecoul B, Chirac P, Trouiller P, Pinle J. Access to essential drugs in poor countries: a lost battle? *J Am Med Assoc* 1999; **281**: 361–7.
13. Shakoor O, Taylor RB, Behrens RH. Assessment of the incidence of substandard drugs in developing countries. *Trop Med Int Health* 1997; **2**: 839–45.
14. Nazerali H, Hogerzeil H. The quality and stability of essential drugs in rural Zimbabwe: controlled longitudinal study. *Br Med J* 1998; **317**: 512–3.
15. Zaffran M. Vaccine transport and storage: environmental challenges. *Dev Biol Stand* 1996; **87**: 9–17.
16. Shoub BD, Cameron NA. Problems encountered in the delivery and storage of OPV in an African country. *Dev Biol Stand* 1996; **87**: 27–32.
17. Grasso M, Ripabelli G, Sammarco ML, Manfredi Selvaggi TM, Quaranta A. Vaccine storage in the community: a study in central Italy. *Bull WHO* 1999; **77**: 352–5.
18. Guthridge SL, Miller NC. Cold chain in a hot climate. *Aust NZ J Public Health* 1996; **20**: 657–60.
19. Jeremijenko A, Kelly H, Sibthorpe B, Attewell R. Improving vaccine storage in general practice refrigerators. *Br Med J* 1996; **312**: 1651–2.
20. Wawryk A, Mavromatis C, Gold M. Electronic monitoring of vaccine cold chain in a metropolitan area. *Br Med J* 1997; **315**: 518–9.
21. Reimer RF, Lewis PR. Vaccine storage in pharmacies on the Central Coast of New South Wales. *Aust NZ J Public Health* 1998; **22**: 274–5.
22. Hanjeet K, Lye MS, Sinniah M, Schnur A. Evaluation of cold chain monitoring in Kelantan, Malaysia. *Bull WHO* 1996; **74**: 391–7.
23. John TJ. Vaccine stability in the context of vaccine delivery in a developing country: India. *Dev Biol Stand* 1996; **87**: 19–25.
24. Adu FD, Adedeji AA, Esan JS, Odusanya OG. Live viral vaccine potency: an index for assessing the cold chain system. *Public Health* 1996; **110**: 325–30.
25. Hogerzeil HV, Battersby A, Srdanovic V, Stjernstrom NE. Stability of essential drugs during shipment to the tropics. *Br Med J* 1992; **304**: 210–2.
26. Kanji N, Hardon A, Harnmeijer JW, Mamdani M, Walt G. *Drugs Policy in Developing Countries*. Atlantic Highlands, NJ: Zed Books, 1992.
27. Ballance R, Pogany J, Forstner H. *The World's Pharmaceutical Industries*. Aldershot: Elgar, 1992.
28. Habiymbere V, Wertheimer AI. Essential drugs should be accessible to all people. *World Health Forum* 1993; **14**: 140–4.
29. Sri-Nyernyaung L. Drug abundance: situation of drugs and drug distribution in the villages of rural

- Thailand. *Asia Pacif J Public Health* 1996-7; **9**: 18-23.
30. Foster S. Supply and use of essential drugs in Sub-Saharan Africa: Some issues and possible solutions. *Soc Sci Med* 1991; **32**: 1201-18.
 31. Kandela P. Shortages distort the social fabric of Iraq. *Lancet* 1998; **351**: 1711-2.
 32. Kent D, Glatzer M. Inappropriate drug-donation practices in Bosnia and Herzegovina, letter. *New Engl J Med* 1998; **338**: 1472-4.
 33. Soumerai SB, Ross-Degnan D. Inadequate prescription-drug coverage for Medicare enrollees—a call for action. *New Engl J Med* 1999; **340**: 722-8.
 34. Soumerai SB, McLaughlin TJ, Ross-Degnan D, Casteris CS, Bollini P. Effects of limiting Medicaid drug-reimbursement benefits on the use of psychotropic agents and acute mental health services by patients with schizophrenia. *New Engl J Med* 1994; **331**: 650-5.
 35. Rabon PG, Linette DC, Gonzalez MF, Garrison S, McGee KH. Limited availability of medications for cancer patients. *South Med J* 1993; **86**: 914-8.
 36. Wise R, Hart T, Cars O, Streulens M, Helmuth R *et al*. Antimicrobial resistance is a major threat to public health [editorial]. *Br Med J* 1998; **317**: 609-10.
 37. Trostle J. Inappropriate distribution of medicines by professionals in developing countries. *Soc Sci Med* 1996; **42**: 1117-20.
 38. Rodruguez-Noriega E, Morfin-Otero R, Esparza-Ahumada S. The use of quinolones in developing countries. *Drugs* 1993; **45**: 42-5.
 39. Ferraz MB, Pereira RB, A Paiva JG, Atra E, Dos Santos JQ. Availability of over-the-counter drugs for arthritis in Sao Paulo, Brazil. *Soc Sci Med* 1996; **42**: 1129-31.
 40. Elzubier AG, Al-Shehry SZ. Too many vitamins for diabetics. *World Health Forum* 1997; **18**: 73-4.
 41. Hogerzeil H, Bimo, Ross-Degnan D, Laing RO, Ofori-Adjei D *et al*. Field tests for rational drug use in developing countries. *Lancet* 1993; **342**: 1408-10.
 42. Gurwitz JH, Avorn J. The ambiguous relation between aging and adverse drug reactions. *Ann Intern Med* 1991; **114**: 956-66.
 43. Hogan DB. Revisiting the O complex: urinary incontinence, delirium and polypharmacy in elderly patients. *Can Med Assoc J* 1997; **157**: 1071-7.
 44. Linzer M, Yang EH, Estes NA, Wang P, Vorperian VR, Kapoor WN. Diagnosing syncope. Part 1: value of history, physical examination and electrocardiography. *Ann Intern Med* 1997; **126**: 989-96.
 45. Stewart RB, Cooper JW. Polypharmacy in the aged. Practical solutions. *Drugs Aging* 1994; **4**: 449-61.
 46. Furniss L, Craig SK, Burns A. Medication use in nursing homes for elderly people. *Int J of Geriatr Psychiatry* 1998; **13**: 433-9.
 47. Soumerai SB, McLaughlin TJ, Spiegelman D, Hertzmark E, Thibault G, Goldman L. Adverse outcomes of under use of β -blockers in elderly survivors of acute myocardial infarction. *J Am Med Assoc* 1997; **277**: 115-21.
 48. Moser M. Why are physicians not prescribing diuretics more frequently in the management of hypertension? *J Am Med Assoc* 1998; **297**: 1813-6.
 49. Rochon PA, Gurwitz JH. Prescribing for seniors. Neither too much nor too little. *J Am Med Assoc* 1999; **282**: 113-5.
 50. Hogerzeil HV, Couper MR, Gray R. Guidelines for drug donations. *Br Med J* 1997; **314**: 737-40.
 51. WHO/Action Program on Essential Drugs. *Globalization and Access to Drugs. Perspectives on the WTO/TRIPS Agreement*. Geneva: WHO, 1999.
 52. Helling-Borda M. The role and experience of the World Health Organisation in assisting countries to develop and implement medicinal drug policies. *Aust Prescriber* 1997; **20**: 34-8.
 53. Ross-Degnan D, Laing R. Improving use of medicines by health providers in primary care. *International Conference on Improving Use of Medicines (ICIUM)*, Chang Mai, 1997.
 54. WHO. 1995. *Report of the WHO Expert Committee on National Drug Policies*. Geneva: WHO.
 55. WHO. *Effective Drug Regulation: What Can Countries Do?* WHO working paper. Geneva: WHO, 1999.
 56. Hogerzeil HV, Walker GJ, Sallami AO, Fernando G. Impact of an essential drugs program on availability and rational use of drugs. *Lancet* 1989; **1**: 141-2.
 57. Molina-Salzar RE, Rivas-Vilchis JE. Overpricing and affordability of drugs: the case of essential drugs in Mexico. *Cadernos Saude Publica* 1998; **14**: 501-16.
 58. Quick J. Affordability and availability of drugs in the private market: promising experiences and cautions. *Aust Prescriber* 1997; **20**: 99-100.
 59. Dumoulin J, Kaddar M, Velásquez G. *Guide to Drug Financing Mechanisms*. Geneva: World Health Organisation, 1998.
 60. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; **342**: 1317-22.
 61. Bero LA, Grilli R, Grimshaw JM *et al*. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *Br Med J* 1998; **317**: 465-8.
 62. Fairley CK, McNeill JJ, Desmond P *et al*. Risk factors for development of flucloxacillin associated jaundice. *Br Med J* 1993; **307**: 1179.
 63. Henry D. Economic analysis as an aid to subsidisation decisions. *Pharmacoeconomics* 1992; **1**: 54-67.
 64. Sketris IS, Hill S. The Australian national publicly subsidized Pharmaceutical Benefits Scheme: any

- lessons for Canada? *Can J Clin Pharmacol* 1998; **5**: 111–8.
65. Commonwealth of Australia. *Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee, Including Major Submissions Involving Economic Analyses*. Canberra: Department of Human Services and Health, 1995.
66. Hill S, Mitchell AM, Henry D. Problems with the interpretation of pharmacoeconomic analyses. *J Am Med Assoc* submitted.

28

Premarketing Applications of Pharmacoepidemiology

HARRY A. GUESS

Merck Research Laboratories, West Point, PA, and University of North Carolina, Chapel Hill, NC, USA

INTRODUCTION

While most of the interest in pharmacoepidemiology has centered on its role in the evaluation of drug safety after marketing,¹⁻³ epidemiologic methods are also useful in evaluating safety concerns that arise prior to market approval. Much of the premarketing patient exposure to many pharmaceutical products occurs in studies conducted without a control group. Often these studies are open extensions of randomized trials in which all patients are offered the option of receiving active treatment upon completion of the blinded phase of the studies. Other sources of premarketing patient exposure without a control group are Treatment IND (Investigational New Drug) Studies⁴ and Compassionate Use Programs.⁵ These programs are intended to give physicians the opportunity to use investigational drugs to treat patients with life-threatening illnesses which are not adequately controlled with approved drugs and for which the investigational drugs are considered likely to be beneficial. Typically, such patients are more seriously ill than

those in the randomized clinical trials designed to evaluate efficacy. Use of epidemiologic methods can sometimes be useful in determining whether rare adverse events in noncomparative clinical studies are occurring at rates above those expected among similar patients not exposed to the study drug. In addition, retrospective analyses of non-comparative clinical trials can help identify risk factors for specific adverse events and thereby contribute to safer use of the study drug.

Premarketing applications of pharmacoepidemiology are similar in many ways to analyses of postmarketing surveillance studies (phase 4 studies). Two important ways in which premarketing and postmarketing applications of pharmacoepidemiology differ are that (i) safety questions arising in clinical trials typically require answers in a matter of days rather than weeks or months, and (ii) the threshold for either the manufacturer or a regulatory agency to decide to halt human exposure to a drug is much lower before market approval than after approval.

In this chapter, we will first review the clinical and regulatory context for epidemiologic evalua-

tions of drug safety during premarketing clinical studies. Next, we will briefly mention information technology requirements and statistical methods. Finally, we will discuss some interesting case studies in which the analyses resulted in publications. Because premarketing applications of pharmacoepidemiology normally require access to premarketing, as yet unpublished, clinical data, such studies are of greatest importance and interest to those in the pharmaceutical industry or regulatory agencies, and the perspective taken in this chapter reflects this. However, the topic should also be of interest to others in the field of pharmacoepidemiology, including academic researchers and clinical investigators.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

The need for premarketing epidemiologic assessments of drug safety may arise in several situations, including evaluation of serious adverse events in noncomparative clinical trials, preparing integrated summaries of safety for a New Drug Application (NDA) or international marketing application, and answering inquiries from regulatory agencies or advisory committees. For clinical trials conducted under US Food and Drug Administration (FDA) regulations for Investigational New Drugs (IND), the Code of Federal Regulations (21 CFR 312.32) currently (January 1999) mandates reporting to FDA and all participating investigators within 15 calendar days any adverse experience which simultaneously meets the three conditions: (i) "serious," (ii) "unexpected," and (iii) having a reasonable possibility of having been caused by the study drug.⁶ If, in addition to the above three conditions, the event is also fatal or life threatening, a report to FDA by telephone or facsimile transmission within seven calendar days is required (6). The publication of these regulations in the Federal Register⁶ noted that they were issued to implement definitions, reporting periods, formats, and standards as recommended by the International Conference on

Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)^{7,8} and by the World Health Organization's Council for International Organizations of Medical Sciences (CIOMS).^{9,10} This is part of an ongoing process of international standardization of definitions and reporting procedures used in drug safety evaluation.

According to current regulations,⁶ a "serious" adverse drug experience is any adverse drug experience occurring at any dose that results in any of the following outcomes: death, a life-threatening adverse drug experience, inpatient hospitalization or prolongation of existing hospitalization, a persistent or significant disability/incapacity, or a congenital anomaly/birth defect. Important medical events that may not result in death, be life threatening, or require hospitalization may be considered a serious adverse drug experience 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.

An "unexpected" adverse drug experience is any adverse drug experience, the specificity or severity of which is not consistent with the current investigator brochure; or, if an investigator brochure is not required or available, the specificity or severity of which is not consistent with the risk information described in the general investigational plan or elsewhere in the current application, as amended. In reporting IND safety reports, the regulations indicate that the sponsor shall identify all safety reports previously filed with the IND concerning a similar adverse experience and shall analyze the significance of the adverse experience in light of the previous similar reports.

An epidemiologic approach to the preparation of an IND safety report can occasionally be useful when the report concerns a rare adverse event for which appropriate comparative incidence data are available. For example cancer incidence and

mortality data by age, gender, and race are available online from a number of countries.^{11–13} To produce an epidemiologic assessment it is necessary to perform, review, and report the analysis in time to meet the 15 calendar day regulatory reporting deadline. This naturally puts a constraint on how comprehensive such an assessment can be.

The FDA has made available a number of guidelines to assist manufacturers in preparing NDAs and FDA medical officers in reviewing them. A recent draft guideline for medical reviewers outlines the goals of a safety review of an NDA:¹⁴

In general, the goals of a safety review are (1) to identify important adverse events that are causally related to the use of the drug, (2) to estimate incidence for those events, and (3) to identify factors that predict the occurrence of those events, e.g., patient factors such as age, gender, race, comorbid illnesses, and drug factors such as dose, plasma level, duration of exposure, concomitant medications. The safety review is useful not only in making a risk/benefit decision, but also in drafting labeling for a drug that is going to be approved.

The guideline also emphasizes the difference between the formal, prespecified efficacy evaluation and the more exploratory approaches used in safety evaluation in pre-approval clinical trials:

Approaches useful for evaluating the safety of a drug under development generally differ substantially from those useful in evaluating its effectiveness. In fact, most of the studies in phases 2–3 of a development program focus on establishing a drug's effectiveness. In designing these trials, critical efficacy endpoints are identified in advance and sample sizes are estimated to permit an adequate test of the null hypothesis. For the most part, phase 2–3 trials are not designed to test hypotheses about safety. In fact, the safety endpoints are generally not known prior to the conduct of these trials, and for many of the observed safety outcomes, one can assume that the available studies are underpowered. The usual approach is to screen broadly and sensitively for adverse events, and it is hoped that this approach should reveal the common adverse event profile of a new drug and detect some of the less common and more serious adverse events associated with

the drug as well. While hypothesis testing for effectiveness outcomes generally is done within individual studies, it usually is not appropriate to proceed with hypothesis testing procedures for safety outcomes. Rather, the approach to safety data may be viewed more as exploration and estimation.

This discussion of the goals and approaches of a pre-approval safety evaluation makes it clear that epidemiologic reasoning and methods are clearly useful. The need to evaluate data from all trials, with different kinds of patient population, differing designs, and differing durations makes epidemiologic approaches especially appropriate. The approach required is one that recognizes the limited nature of the data and avoids overinterpretation in either direction.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY RESEARCH

EPIDEMIOLOGIC APPROACHES TO ASSESSING CAUSALITY OF ADVERSE EVENTS IN PRE-APPROVAL CLINICAL TRIALS

Clinical investigators and clinical monitors continually review reports of adverse events occurring in pre-approval clinical trials. Investigators are asked to provide a clinical assessment as to the causality of each event. Such assessments are a necessary part of safety monitoring in clinical research, although they are known to be subjective and imprecise¹⁴ (see also Chapter 32). Criteria to help guide causality assessments have been published by FDA.¹⁵ These criteria are most useful when there are well defined differences in clinical features of drug related and non-drug-related cases of the adverse event. However, for serious, uncommon adverse events where the clinical features of the drug related cases could be similar to those of non-drug-related cases, it can sometimes be helpful to supplement clinical causality assessments of individual cases by epidemiologic

assessment criteria based on comparisons between groups of patients.

The epidemiologic literature provides several sets of criteria for helping to decide whether an empirical association is likely to be causal. The best known criteria are those proposed in 1965 by Bradford-Hill to help evaluate evidence linking cigarette smoking with lung cancer¹⁶ (see also Chapter 2). The nine Bradford-Hill criteria are discussed briefly below as they relate to evaluation of adverse experiences in premarketing clinical trials.

Strength of Association

This is commonly quantified in terms of a suitably adjusted hazard function ratio or risk ratio (relative risk), rather than a *p*-value. In general, the farther the ratio is from unity, the less likely it could be entirely attributable to imbalances in risk factors between groups. An exception to this occurs when the ratio is based on very small numbers or highly influenced by only a few cases.

In addition, it is worth noting that with the large number of different kinds of adverse events often seen in large clinical trials, it is quite likely that some risk ratios far from unity will occur by chance alone. Multiple comparisons not only distort *p*-values, but can also bias risk ratios and their confidence intervals away from unity. This happens because screening many different adverse event terms and selecting those with very low *p*-values inherently selects for relative risks that are biased away from unity.¹⁷ Since the confidence interval is centered about the observed relative risk, it will also be biased away from unity. This type of selection bias, a form of regression to the mean, is common to all programs of screening for unusually large or small values and is well known in the statistics literature. One example where this appears to have occurred is in some observational studies of vasectomy and prostate cancer.^{17,18} Because pre-approval reviews of drug safety use the same set of data for identifying potentially drug related events and for providing preliminary estimates of the risk, the relative risk estimates can be biased away from unity.

It is also important to recognize that the *absence* of any association between a drug and any given adverse event has to be judged in the context of the limited amount of patient exposure in pre-approval clinical trials. This topic is reviewed in ICH Guideline E1A, which addresses the extent of population exposure needed to assess clinical safety for drugs intended for long term treatment of non-life-threatening conditions.^{7,8} The guideline notes that

It is expected that short-term event rates (cumulative 3-month incidence of about 1%) will be well characterized. Events where the rate of occurrence changes over a longer period of time may need to be characterized depending on their severity and importance to the risk–benefit assessment of the drug. The safety evaluation during clinical drug development is not expected to characterize rare adverse events, for example, those occurring in less than 1 in 1000 patients.

Consistency

The original wording was, “Has [the association] been repeatedly observed by different persons, in different places, circumstances and times.” This is a useful criterion for assessing adverse experiences in a program of several clinical trials. Results that show a consistently elevated incidence on drug across studies are generally more convincing than those in which the elevated risk is largely due to one study. It has been noted by FDA, however, that an apparent lack of consistency among trials may simply reflect differences in trial design, making the event less likely in some trials than others. For example, in a combined analysis of several studies in an NDA, a reassuringly low incidence of phototoxicity was seen. Examination of individual studies revealed that in one study the rate of phototoxicity was substantial—a finding that had been obscured by combining the data. This study was the only outpatient study and thus these were the only patients who had an opportunity to have the event by virtue of having sun exposure.¹⁴

Temporality

In both epidemiologic and clinical assessments of causality, it is important to distinguish between

events having onset before a drug was started and those having onset after drug therapy was started. Especially in the evaluation of adverse experiences from studies without a comparison group, it sometimes occurs that early symptoms of a disease which is present but not yet recognized lead a patient to be prescribed a drug, which then appears to be the cause of the disease when it is eventually diagnosed. This has been called “protopathic bias” and is a special case of the broader concept of “confounding by indication” (see also Chapter 34). A classic example of this form of bias was seen in uncontrolled postmarketing surveillance studies of cimetidine, where a higher than expected incidence of gastric carcinoma was found among users than among nonusers. It is likely that many of the cancers were present but undiagnosed at the time the cimetidine was started. Subsequent studies showed that elevations in gastric cancer risk diminished with duration of followup, returning to baseline with long term use.¹⁹

Another way in which the concept of temporality plays a role in the evaluation of adverse experiences is whether the timing of the reaction in relation to duration of exposure is consistent with the proposed mechanism. Thus, an elevated incidence of cancer in patients who had been taking a drug for several years would be of more concern than would an elevated incidence in the first year of therapy. Timing plays a major role in the evaluation of adverse events that are thought to be due to immune mechanisms (e.g., anaphylaxis, angioedema, hemo-

lytic anemia, serum-sickness), altered metabolism, or drug interactions.¹⁵ A classic example where timing provided highly convincing evidence of causality was in the evaluation of Guillain-Barré syndrome in association with the “swine-flu” vaccine²⁰ (see Figure 28.1).

Dose–Response

Adverse effects caused by exaggerated pharmacological actions of a drug are often dose dependent. Examples include hypotension resulting from the use of antihypertensive drugs and gastrointestinal hemorrhage from nonsteroidal anti-inflammatory drugs. For such adverse events it is especially important to characterize how the incidence of the event varies with dose in different patient populations. This is not only important in assessing causality and quantifying incidence, but also in understanding the mechanism and providing guidance to clinicians. The case of renal failure in patients with congestive heart failure treated with the angiotensin converting enzyme (ACE) inhibitor enalapril provides an excellent example of where a life saving drug was initially thought by some commentators to be inherently nephrotoxic, when the actual problem was that the starting dose in patients with severe congestive heart failure was too high. While ACE inhibitors improve survival in patients with congestive heart failure, too high a dose can shut off renal function, because angiotensin-II is part of the compensatory process for

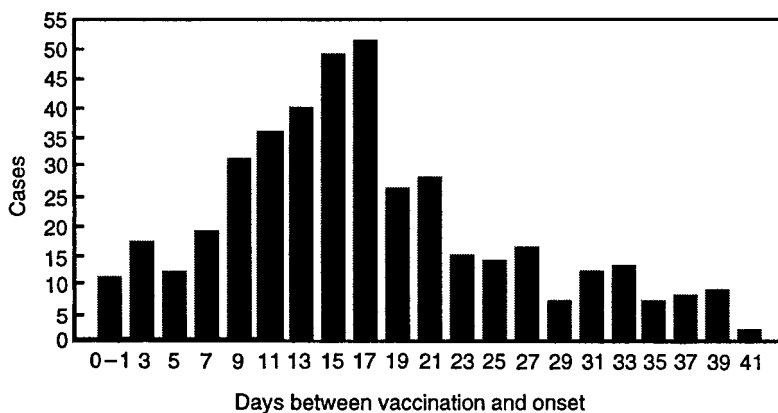


Figure 28.1. Onset of Guillain-Barré syndrome in relation to day of vaccination.¹⁶ Reproduced with permission.

maintaining adequate renal perfusion pressure in the face of low cardiac output. This example illustrates the importance of understanding both pathophysiology and dose–response relationships in evaluating drug safety.

Experimental Evidence

Evidence from animal experiments or from previous human clinical trials can be especially important in helping to interpret adverse experiences. Well designed animal studies can be especially helpful in determining the extent to which animal results are applicable to humans. On the other hand, poorly designed animal studies may produce misleading results, which can require carefully designed further experiments to correct.

Biological Plausibility

In the original statement of biological plausibility as one of the considerations in judging causality, Bradford-Hill noted

It will be helpful if the causation we suspect is biologically plausible. But this is a feature I am convinced we cannot demand. What is biologically plausible depends on the biological knowledge of the day.

He went on to note, among other examples, that the role of rubella in causing congenital malformations was initially doubted on the basis of presumed lack of biological plausibility.¹⁶ As important as we consider biological plausibility to be, it is equally important to realize that it can mislead in either direction. It is worth noting, however, that some epidemiologists have expressed the view that too little attention is currently paid to biological plausibility in reviews of epidemiologic evidence.²¹

Coherence

The postulated cause-and-effect interpretation should not seriously conflict with what is known of the natural history and biology of the event. In this regard, biological and laboratory evidence

may strengthen a causal interpretation, but lack of it cannot be used to nullify one.

Analogy

Reasoning by analogy often serves as a basis for having a lower threshold for judging an adverse event to be causally related to one drug in a class when the same effect is considered to be causally related to another drug in the class. Such reasoning is even more problematic than reasoning based on biological plausibility and this criterion has been criticized as being potentially misleading. It is worth noting, however, that certain types of adverse event are commonly enough associated with drugs to be worth anticipating as topics which regulatory agencies are likely to ask to be specifically addressed. The FDA medical reviewer guidance document¹⁴ states that

There is an expanding checklist of adverse events that we now routinely think about when reviewing new drugs, e.g., sedation, anticholinergic effects, vasodilator effects, QT prolongation, tachycardia, beta agonist effects, hepatotoxicity, hematological effects, such as neutropenia. All of these standard concerns should be specifically addressed within the appropriate body systems.

The document suggests that medical reviewers perform a “review of body systems” to help judge whether potential safety concerns within each body system had been adequately evaluated¹⁴

...within each body system, the reviewer should comment on the adequacy of the development program in evaluating the new drug with regard to the body system in question. This gets at the question of whether or not “all tests reasonably applicable” were conducted to assess the safety of the new drug. Were all the appropriate animal tests done? Were all the appropriate clinical tests done? Were all potentially important findings adequately explored, e.g., to what extent was psychomotor impairment specifically assessed in a drug that is sedating? If not, this could be the basis for a non-approval action, or alternatively, a phase 4 study for additional testing....

Thus, it is important to anticipate potential mechanism based toxicity and to design the drug

development program to address concerns that can reasonably be anticipated and tested. This requires understanding both the pharmacology of the compound and its class and the natural history and epidemiology of the disease being treated.

Specificity

The finding that an adverse event has a very specific presentation or is associated with a specific histopathology can be useful evidence in favor of causality. Absence of specificity in an epidemiologic study often manifests itself as elevation in positive associations between an exposure and a large number of unrelated outcomes. Such a finding raises the possibility of uncontrolled confounding or selection bias.

In summary, while the Bradford-Hill criteria were originally proposed for use in interpreting evidence from observational studies, they also provide a good framework for evaluating evidence from unplanned comparisons of adverse experiences in pre-approval clinical trials.

Another approach to assessing causality is through Bayesian causality assessment,^{22,23} which provides a framework for using clinical and epidemiologic information to compute a probability that a specific drug caused a specific set of events in a specific patient. In Bayesian causality assessments the statement that a given drug “D” caused an event “E” is defined to mean that the event would not have happened as and when it did, if drug “D” had not been administered. Using this definition, a Bayesian causality assessment begins by decomposing the calculation of a probability of causation into a number of subcalculations, some of which make use of epidemiologic information and others of which make use of pharmacologic and medical information specific to the case. Thus, rather than being alternatives to epidemiologic assessments, Bayesian causality assessments *require* epidemiologic assessments as a (somewhat hidden) part of a process which yields a numerical probability of causation. Bayesian analysis is discussed in more detail in Chapter 32.

Statistical Analysis Methods

While both cohort and case-control study designs are commonly used in postmarketing pharmacoepidemiology, premarketing studies typically involve only cohort designs. One of the most common problems in premarketing pharmacoepidemiology is to compare the incidence of a given adverse event in the cohort of patients exposed to the study drug to the incidence in a suitably chosen cohort of historical controls. Here *incidence* is used in the epidemiologic sense, where the numerator refers to the number of events (counting only the initial event in each patient) and the denominator often refers to the total *person time at risk* during exposure to the study drug, sometimes referred to as incidence density. Person time at risk is typically measured in *person days*, where each patient contributes one person day for each day he or she is on the study drug and is at risk for the given adverse event. The person time of exposure for each patient is thus most commonly measured from the day of first exposure to the study drug, up to the earlier of (i) the date of last followup for the patient or (ii) the day that the patient first experienced the adverse event. When analyzing incidence it is important to review the data for any evidence of temporal trends in relation to start of therapy. When such trends are found, it may be necessary to produce separate risk computations for different windows of time on therapy. This is part of the general topic of effect modification by time on therapy, as discussed elsewhere²⁴ (see also Chapter 43).

For example, Suissa and colleagues recently showed that when risk in relation to duration of therapy was taken into account in comparing the risk of venous thromboembolism among new users of second and third generation oral contraceptives, the incidences in the two groups were much closer than had been found in previous analyses where this approach had not been used.²⁵ This phenomenon has been termed “effect modification by duration of therapy”²⁴ or “depletion of susceptibles”.²⁶

Analyses of events in relation to person time can be analyzed by Cox regression, Poisson regression,²⁷ or by stratified incidence density

computations²⁸ using standard software packages, such as Stata[®] or SAS[®]. When the sample sizes are very small confidence intervals on the stratified relative risk estimates may be computed by exact methods.²⁹ Analytic methods and their limitations are further discussed in standard references.^{28,30} Issues that often arise include calculation of point and interval estimates of standardized morbidity or mortality ratios (SMRs) (30, p 65), comparing adjusted risk ratios for different groups of patients (30, pp 106–109), and testing for heterogeneity (representing potential effect modification) and dose–response trend (30, p 110). As with all epidemiologic studies, it is essential to approach these analyses with recognition of the biological and clinical aspects of the problem, as well as with an understanding of the meaning and limitations of the formal computational methods. Examples of applications of these methods are given in a later section.

Since pre-marketing epidemiologic evaluations typically have to be performed under tight time constraints, it is sometimes necessary to use published data to help estimate risks. One method which has been used in environmental epidemiology, but does not appear to have been used yet in pharmacoepidemiology, involves using risk factor prevalence in cases together with relative risks from published studies to produce adjusted incidence ratios.³¹ Using this method, one can compute risk comparisons with an external control group in which the prevalence of risk factors differs from those in the cases that have occurred in the clinical study.

Dean and colleagues³¹ presented this method and used it to determine whether the observed elevation in breast cancer incidence in one community with a contaminated water supply was higher than in surrounding communities, after adjusting for differences in prevalence of known risk factors for breast cancer. The method of adjustment makes use of the population attributable risk fraction (AR) for each population, i.e., the fraction of risk in each population attributable to the known risk factors (32, p 163). The complementary fraction, $1 - AR$, is the fraction of risk that is *not* attributable to the known risk factors. If I_E is the total incidence in the exposed community, then $I_E(1 - AR)_E$ is the incidence of

breast cancer in the exposed community that is *not* attributable to the known risk factors. The corresponding quantity, $I_C(1 - AR)_C$, is the incidence of breast cancer in the comparison communities that is *not* attributable to the known risk factors. Hence, the ratio of these two quantities is the incidence ratio for breast cancer *not* attributable to the known risk factors. This is the risk ratio of interest for comparing risk in the two communities, while adjusting for the difference in prevalence of risk factors.

The novelty of the approach lies in using the prevalence of risk factors in only the cases to estimate the attributable risk fraction, AR. The method for doing this was outlined by Bruzzi and colleagues (33, equation (6)) and involves using published information on relative risks associated with known risk factors to compute the values of AR for the exposed community and the control community. Here information from published sources is used to compute a relative risk for each case based on all known risk factors for that case, relative to some reference group. Of course, the same reference group and the relative risk information must be used for the cases in the exposed community and the cases in the control communities. One major drawback of this approach is that no methods for expressing the statistical uncertainty of the resulting incidence ratio have been published.

This approach would appear to be mainly useful in noncomparative drug studies to provide a rough idea whether what appeared to be an elevated incidence of an adverse event could be attributed to a greater prevalence of risk factors in the exposed group, relative to the prevalence in some historical control group. For example, the method would appear to be of some value in evaluating adverse events among a group of very ill patients in a compassionate use program. The advantage of only requiring risk factor data from the cases is that sometimes these are the only patients for whom information on important risk factors for the adverse events is available.

Information Technology Requirements

Effective premarketing pharmacoepidemiology requires considerable advanced planning to make

sure that information can be assembled, analyzed, summarized, reviewed, and reported in hours or days. For historical data to be useful, it must be either available in sufficient detail from published sources or from existing datasets that can be accessed, analyzed, and checked for errors within a day or two. This can only be done if one can anticipate at least some types of problems likely to arise and have appropriate sources of epidemiologic information readily accessible.

In addition, the clinical trials data management organization has to be able to produce accurate, suitably detailed patient exposure information from many different trials in a number of different countries. This is more difficult than it might initially appear to be. Because of regulatory reporting requirements, adverse experience information (the numerator) will typically be current to within a few days, while the patient exposure information (the denominator) may be weeks behind and may not yet have essential covariate information and demographics in accurately retrievable form.

Getting new drugs approved in a timely manner can depend on the ability to answer safety questions rapidly and accurately, and that can depend on the way in which data management is staffed, organized, and equipped. One of the rate-limiting factors in securing timely approval is the manpower needed to process worldwide clinical trial data on thousands of patients from many different countries accurately. This entails producing both standard efficacy and safety analyses and special *ad hoc* safety analyses required for responding to questions raised by reviewers at drug regulatory agencies. The data management problems posed by the need to answer *ad hoc* safety questions rapidly are different from those posed by the need for the formal statistical analyses of efficacy required for drug approval. To be successful, the data management organization and systems must meet both challenges.

As potential capabilities of data management systems increase, the potential benefit from using them effectively increases dramatically, as does the potential gap in productivity from using them ineffectively.³⁴ For example, a request by a regulatory agency for an analysis of certain types of laboratory data on all patients exposed to a

study drug for more than two months might take only a day or two for a response if the data from all worldwide studies were stored so that a query could readily be run against the entire database. However, the query could take weeks and consume much more scarce manpower critically needed for other tasks if: the specific data requested were stored differently by different countries; the way the data were stored made custom programming necessary; a decision had been made not to have centralized computer storage of the specific data needed; the laboratory methods used in different countries were different and cross-translation files had not been centrally stored; there were problems with getting data entered, checked, and available for analysis in a timely manner; or there were other organizational impediments to producing a unified worldwide tabulation and analysis. During the course of regulatory review of a new drug many standard and *ad hoc* analyses must be performed for different regulatory agencies in different countries. The difference between effective and somewhat less effective clinical data management and analysis can easily translate into differences of several weeks in drug approval time. Such differences can translate into substantial lost revenue for each new drug.

CURRENTLY AVAILABLE SOLUTIONS

Epidemiologic investigation of pre-marketing drug safety problems is a collaborative undertaking that requires the ability not only to respond to inquires promptly and accurately, but also to recognize and answer questions that should have been asked but were not. Sometimes the initial request from the clinical or regulatory group in a company to the epidemiology group asks for a specific tabulation or calculation, which may not in fact represent the most appropriate way to approach the problem. Requests transmitted through several intermediaries often change in meaning, so that what finally reaches the person responsible for the analysis may be quite different from what was originally asked. When the original request is a written inquiry (e.g., from a regulatory agency), it is important to read the actual written question and the surrounding

context before attempting to answer a paraphrased version of it.

As with all epidemiologic and statistical consulting, the best way to approach a request for consultation on a potential safety problem with a premarketing investigational drug is to first understand the broader context surrounding the immediate inquiry. This includes gaining at least a basic understanding of the pharmacology, mechanism of action, preclinical toxicity profile, and clinical safety profile of the compound. It also entails understanding the characteristics and comorbidities of the patient population in which the drug has been studied. Finally, it requires knowing how and in what context the question arose. A review of several case studies may be helpful in conveying these concepts.

One example, which illustrates the interplay between clinical pharmacology and epidemiology, involved seizures in seriously ill hospitalized patients with systemic gram-negative infections being treated with the β -lactam antibiotic, imipenem/cilastatin. During initial randomized clinical trials few seizures were reported, either with imipenem/cilastatin or with control antibiotics. In subsequent noncomparative studies an increasing number of seizures were noted, often in patients with predisposing factors, such as compromised renal function that could alter drug metabolism and affect serum levels of the antibiotic. An association of seizure risk with antibiotic level was biologically plausible, in light of known epileptogenic properties of β -lactam antibiotics. At the time these studies were conducted, laboratory techniques to measure serum levels had not yet become widely available and so serum levels on these patients were rarely known.

To study seizure risk in relation to serum level, clinical pharmacology studies were reviewed to determine how serum levels varied with dose, body weight, gender, age, and renal function. An equation was developed to predict serum level as a function of these parameters. The predicted serum levels were then used as one variable in an analysis of seizure risk in the noncomparative clinical trial patients. It was found that risk of seizure was strongly and independently related to predicted serum level, after controlling for several

non-drug-related seizure risk factors.³⁵ At the same time, however, the other factors found to be related to seizure risk in patients receiving imipenem/cilastatin were also found to be risk factors for seizures in patients who had not received imipenem/cilastatin. These factors included a history of seizures, central nervous system insults, and renal impairment.³⁶ Age was not found predictive of seizures when adjustment for the above factors was made.

This study illustrates the concept of *pharmacoe*-epidemiology in a study where methods of clinical pharmacology and epidemiology were both brought into play to investigate and solve a problem that arose during the course of premarketing clinical trials. It also illustrates the important point that merely quantifying risk and identifying patients at increased risk would not have been sufficient. What was needed was to identify measures to help reduce the risk and to help educate physicians on the need for dosage adjustments. The investigation into seizure risk led to improved prescribing information with better recommendations for dosage adjustments in the presence of impaired renal function.

An illustration of statistical methods for pharmacoeepidemiology cohort studies is provided by a study comparing the observed incidence of Guillain-Barré syndrome among recipients of a plasma-derived hepatitis-B vaccine with that which would be expected in the general population, controlling for age in a stratified analysis.³⁷ Although this study involved postmarketing data, it provides a good illustration of the type of cohort study methods typically used in pre-marketing analyses. During the first three years of marketed use of the vaccine, nine cases of Guillain-Barré syndrome were reported among vaccinees. Six of the nine cases occurred within six weeks after the first dose and the remaining three occurred within eight weeks after either the second or third dose. (The vaccine was administered in three injections, with the second occurring one month after the first and the third occurring six months after the first.) Based on knowledge of the temporal pattern of risk of Guillain-Barré syndrome with other vaccines and on the serological response of the hepatitis-B vaccine, a decision was made to

consider four analyses based on two time windows (six weeks or eight weeks) applied to either the first dose only or to all three doses. The total number of vaccinees was estimated from sales data and the age distribution was estimated from surveys. Comparison data were obtained on the age-specific incidence of Guillain-Barré syndrome in several studies, including a community based study of Guillain-Barré syndrome among residents of Olmsted County, MN. Because the number of cases in vaccinees was so small, the confidence intervals about the relative risks were calculated by an exact binomial method.²⁹ The age adjusted relative risks in the four analyses ranged from the lowest value of 1.7 (95% confidence interval: 0.8–3.6), based on nine cases occurring in an at-risk interval of 8 weeks after each of the three doses, to the highest value of 3.6 (95% confidence interval: 1.3–8.3), based on six cases occurring in an at-risk interval of six weeks after the first dose only (37, Table 6). Although the numerator (number of cases) in the latter analysis was less than in the former, the denominator (person time at risk) was also less, and the age adjusted relative risk was higher.

Another example involved the question of whether herpes zoster might be expected to occur more frequently following vaccination with a live-virus varicella (chickenpox) vaccine than following natural varicella. It was found that the incidence and clinical characteristics of herpes zoster in childhood following natural varicella were not well enough known to permit any meaningful comparison with cases of childhood herpes zoster following vaccination with the varicella vaccine. This led to an epidemiologic study to document the incidence and clinical features of herpes zoster following natural varicella.³⁸ It was found that herpes zoster in children was both much more common than had been previously thought and also was much milder than herpes zoster in adults. Using these data as a point of comparison, subsequent studies suggested that the incidence of herpes zoster in childhood following vaccination did not appear to be more common with the vaccine than with natural varicella.³⁹

In this example, what was needed was more information about the incidence and clinical

spectrum of the adverse event following natural varicella. This illustrates the point that to quantify risk associated with a therapeutic or prophylactic intervention, it is often necessary to develop additional information about the natural course of the disease itself.⁴⁰ Such studies need to be started very early in drug or vaccine development in order to be able to have information available when needed to address safety concerns that arise in clinical studies.

A third example evaluated the risk of angioedema in relation to angiotensin converting enzyme inhibitors, based on data from three large studies involving about 12 000 patients each.⁴¹ Previous studies had suggested that angioedema was more common early in therapy than later in therapy. This hypothesis was tested by examining incidence as a function of time on therapy. It was found that the incidence in the first week of therapy was about one case per 3000 patients per week; thereafter the incidence was about 14-fold lower, with no evidence of any temporal trend in incidence beyond the first week of therapy (Figure 28.2).

This study illustrates several important points. First, it is essential to establish a clear case definition in any epidemiologic study. Failure to do this can lead to continued confusion about what is being counted. The epidemiologic case definition may be different from whatever case definition is used for regulatory reporting. (The latter definition is typically maximally inclusive, while the former may be more restrictive.) In this study, the clinical case definition was established in consultation with clinical experts in the diagnosis and treatment of angioedema. Using this case definition, all case reports that might have represented angioedema were re-reviewed, classified, abstracted, and used in computing incidence. Second, the study illustrates the importance of examining risk in relation to time on therapy.^{24–26} Any temporal patterns detected will be of clinical relevance. Further, failure to account for temporal differences can invalidate calculation of relative risks. Finally, this study shows how to combine epidemiologic analyses with a clinical review of case reports so as not only to quantify the incidence and temporal pattern of occurrence of

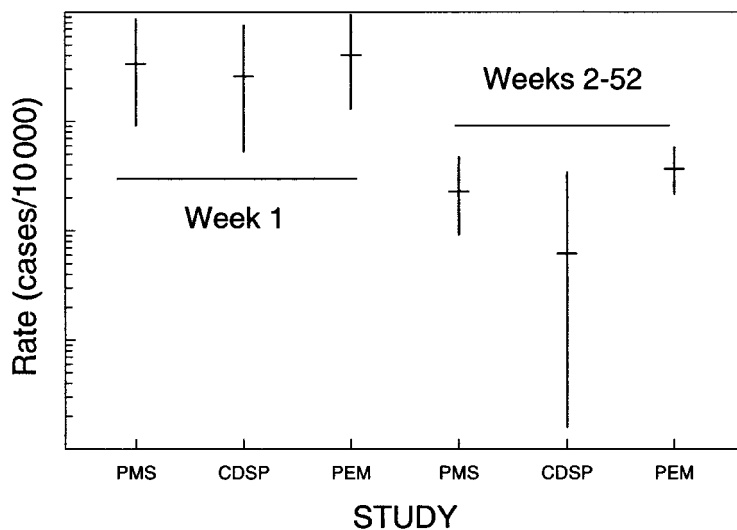


Figure 28.2. Angioedema associated with angiotensin converting enzyme inhibition. Incidence in relation to time on drug. The lines show the 95% confidence intervals on the observed incidence rates. The central bar on each line represents the observed incidence rate. The abbreviations for the three studies on which these rates were based are PMS = postmarketing surveillance study, CDSP = clinical development study program, and PEM = prescription event monitoring study. These are explained in the publication from which these data were taken, with permission.³⁷

an adverse event, but also to describe its clinical features and spectrum of severity.

THE FUTURE

Premarketing and postmarketing applications of pharmacoepidemiology differ in the speed of response required and in the fact that the threshold for halting human exposure to a drug is lower before market approval than after market approval. In addition, premarketing pharmacoepidemiology uses cohort study designs almost exclusively, while postmarketing pharmacoepidemiology often uses case-control study designs as well. Safety monitoring in clinical trials can be improved by anticipating potential questions and organizing data resources in advance to answer them as rapidly as possible. Having a responsive data management system is as essential to premarketing applications of pharmacoepidemiology as it is to the entire process of drug development. Epidemiologic planning should be regarded as one of the first steps in drug development, rather than as something to be considered at or after market

approval. This requires effort to build and maintain a repository of disease incidence and natural history data from governmental surveys, public-use data tapes of governmental studies, publications, and other sources.⁴² It also requires undertaking epidemiologic studies far enough in advance of clinical trials to be able to contribute to clinical trial planning and analysis. Finally, it requires that the epidemiologic group maintain an awareness of pertinent research findings in areas relevant to the diseases under study and to the types of measurement and analytic technique likely to be needed.

REFERENCES

1. Lawson DH. Pharmacoepidemiology: a new discipline. *Br Med J* 1984; **289**: 940-1.
2. Tilson HH. Pharmacoepidemiology: the future. *Drug Intell Clin Pharm* 1988; **22**: 416-21.
3. Guess HA. Adverse and beneficial effects in pharmacoepidemiology. In: Armitage P, Colton T, eds, *Encyclopedia of Biostatistics*. New York: Wiley, 1998; 3332-8.

4. Mattison N. The FDA's treatment IND: current controversies. *Pharmaceut Med* 1988; **3**: 159–71.
5. Calandra GB, Garelik JP, Kohler PT, Brown KB. Problems and benefits of an antibiotic compassionate therapy program. *Rev Infect Dis* 1987; **9**: 1095–101.
6. *Federal Register*. Expedited Safety Reporting Requirements for New Drugs and Biologics. 7 October, 1997, **62** (194): 52 237–53. Available at <http://www.access.gpo.gov/sudocs/aces/aces140.html>.
7. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Available at: <http://www.ifpma.org/ich1.html>. Accessed 30 January, 1999.
8. ICH Topics and Guidelines. Available at: <http://www.ifpma.org/ich5.html>. Accessed 30 January, 1999.
9. Council for International Organizations of Medical Sciences (CIOMS). Available at <http://www.who.int/dsa/cat98/zcioms8.htm>. Accessed 30 January, 1999.
10. CIOMS Working Groups I–IV. Available at: <http://www.who.int/dsa/cat98/zcioms8.htm>. Accessed 30 January, 1999.
11. National Cancer Institute. *Surveillance, Epidemiology and End Results*. Available at: <http://www.seer.ims.nci.nih.gov/>. Accessed 25 November, 1998.
12. Rotterdam Cancer Registry. *Cancer Registries on the Web*. Available at: <http://www.ikr.nl/canregs.htm>. Accessed 25 November, 1998.
13. International Agency for Research on Cancer. *Cancer in the European Union*. Available at: <http://www-dep.iarc.fr/dataava/eucandat.htm>. Accessed 25 November, 1998.
14. US Food and Drug Administration, Center for Drug Evaluation and Research. *Conducting a Safety Review of a New Product Application and Preparing a Report on the Review*. Available at: <http://www.fda.gov/cder/guidance/index.htm>. Accessed 22 November, 1998.
15. Food and Drug Administration Staff College. *Clinical Therapeutics and the Recognition of Drug-Induced Disease*. MedWatch Continuing Education Article, June 1995. Available at: <http://www.fda.gov/medwatch/articles/dig/rcontent.htm>. Accessed 24 November, 1998.
16. Hill AB. The environment and disease: association or causation. *Proc R Soc Med* 1965; **58**: 295–300.
17. Guess HA. Invited commentary: vasectomy and prostate cancer. *Am J Epidemiol* 1990; **132**: 1062–5.
18. Rosenberg L, Palmer JR, Zauber AG, Warshauer ME, Strom BL, Harlap S, Shapiro S. The relation of vasectomy to the risk of cancer. *Am J Epidemiol* 1994; **140**: 431–8.
19. Walker AM. Confounding by indication [editorial]. *Epidemiology* 1996; **7**: 335–6.
20. Schonberger LB, Bregman DJ, Sullivan-Bolyai JZ, Keenlyside RA, Ziegler DW, Retailiau HF, Eddins DL, Bryan JA: Guillain-Barré syndrome following vaccination in the National Influenza Immunization Program, United States, 1976–1977. *Am J Epidemiol* 1979; **110**: 105–23.
21. Weed DL, Hursting SD: Biologic plausibility in causal inference: current method and practice. *Am J Epidemiol* 1998; **147**: 415–25.
22. Lane DA. The Bayesian approach to causality assessment: an introduction. *Drug Inf J* 1986; **20**: 455–61.
23. Naranjo CA, Lanctot KL, Lane DA. The Bayesian differential diagnosis of neutropenia associated with antiarrhythmic agents. *J Clin Pharmacol* 1990; **30**: 1120–7.
24. 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.
25. Suissa S, Blais L, Spitzer WO, Cusson J, Lewis M, Heinemann L. First-time use of newer oral contraceptives and the risk of thromboembolism. *Contraception* 1997; **56**: 141–6.
26. Yola M, Lucien A: Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol* 1994; **47**: 731–7.
27. Frome EL, Checkoway H. Epidemiologic programs for computers and calculators. Use of Poisson regression models in estimating incidence rates and ratios. *Am J Epidemiol* 1985; **121**: 309–23.
28. Greenland S, Rothman KJ. Introduction to stratified analysis. In: Rothman KJ, Greenland S, eds, *Modern Epidemiology*, 2nd edn. Philadelphia, PA: Lippincott–Raven, 1998; Ch. 15, 253–79.
29. Guess HA, Thomas JE. A rapidly converging algorithm for exact binomial confidence intervals about the relative risk in follow-up studies with stratified incidence density data. *Epidemiology* 1990; **1**: 75–7.
30. Breslow NE, Day NE. *Statistical Methods in Cancer Research, Volume II: the Design and Analysis of Cohort Studies*, International Agency for Research on Cancer Scientific Publication No. 82. Lyon: World Health Organization, 1987.
31. Dean AG, Imrey HH, Dusich K, Hall WN. Adjusting morbidity ratios in two communities using risk factor prevalence in cases. *Am J Epidemiol* 1988; **127**: 654–62.
32. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. Belmont, CA: Lifetime Learning, 1982; 163.
33. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol* 1985; **122**: 904–14.
34. Guess HA. Worldwide drug safety information processing: what's ahead. *Drug Inf J* 1989; **23**: 169–70.

35. Calandra G, Lydick E, Carrigan J, Weiss L, Guess H. Factors predisposing to seizures in seriously ill infected patients receiving antibiotics: experience with Imipenem/Cilastatin. *Am J Med* 1988; **84**: 911–8.
36. Guess HA, Resseguie LJ, Melton LJ III, Kurland LT, Lydick EG, Wilson WR. Factors predictive of seizures among intensive care unit patients with gram-negative infections. *Epilepsia* 1990; **31**: 567–73.
37. Shaw FE, Graham DJ, Guess HA, Milstien JB, Johnson JM, Schatz GC, Hadler SC, Kuritsky JN, Hiner EE, Bregman DJ, Maynard JE. Postmarketing surveillance for neurological adverse events reported after hepatitis B vaccination: experience of the first three years. *Am J Epidemiol* 1988; **127**: 337–52.
38. Guess HA, Broughton DD, Melton LJ III, Kurland LT. Epidemiology of herpes zoster in children and adolescents: a population-based study. *Pediatrics* 1985; **76**: 512–7.
39. Plotkin SA. Hell's fire and varicella–vaccine safety [editorial]. *New Engl J Med* 1988; **318**: 573–5.
40. Guess HA, Stephenson WP, Sacks ST, Gardner JS. Beyond pharmacoepidemiology: the larger role of epidemiology in drug development. *J Clin Epidemiol* 1988; **41**: 995–6.
41. Slater EE, Merrill DD, Guess HA, Roylance PJ, Cooper WD, Inman WHW, Ewan PW. Clinical profile of angioedema associated with angiotensin converting-enzyme inhibition. *J Am Med Assoc* 1988; **260**: 967–70.
42. Guess HA. The role of epidemiology in pharmaceutical research, development, and marketing. In: Hartzema AG, Porta MS, Tilson HH, eds, *Pharmacoepidemiology: an Introduction*, 2nd edn. Cincinnati, OH: Whitney, 1991; 321–9.

29

Studies of Drug Utilization

DAVID LEE

Management Sciences for Health, Arlington, VA, USA

ULF BERGMAN

Department of Clinical Pharmacology, Karolinska Institute, Huddinge University Hospital, S-141 86 Stockholm, Sweden

INTRODUCTION

DEFINITIONS

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.”¹ Some authors have suggested that the development of drugs relative to health priorities should also be included.² This broad definition differs from the more narrow one which appeared in the North American literature, “the prescribing, dispensing and ingesting of drugs”.^{3,4}

In both of the above definitions, recognition is granted, explicitly or implicitly, of the nonpharmacologic (socioanthropological, behavioral, and economic) factors influencing drug utilization. Studies of the process of drug utilization focus on the factors influencing and events involved in the prescribing, dispensing, administration, and taking of medication. However, the broader

definition of the WHO goes beyond the “process” or “pharmacokinetic” aspect of drug utilization, that is the movement of drugs along the therapeutic drug chain, to include consideration of the various “outcomes” or “pharmacodynamics” of drug use.⁵ According to this definition, studies of drug utilization include not only studies of the medical and nonmedical aspects influencing drug utilization, but also the effects of drug utilization at all levels. Studies of how drug utilization relates to the effects of drug use, beneficial or adverse, are usually labeled analytic pharmacoepidemiology research. These two aspects of the study of drug utilization have developed along parallel lines, but may now be regarded as interrelated and part of a continuum of interests and methodologies.^{6,7}

As stated by Lunde and Baksaas,⁸ the general objectives of drug utilization studies are “problem identification and problem analysis in relation to importance, causes, and consequences; establishment of a weighted basis for decisions on problem solution; assessment of the effects of the action taken. These objectives are relevant to problems

and decision making throughout the drug and health chain. The approaches may vary according to the purpose and the needs of the users. Those include the health authorities, the drug manufacturers, the academic and clinical health professionals, social scientists, and economists as well as the media and the consumers.”

This chapter focuses on the current status of descriptive epidemiological approaches to the study of the processes (or “pharmacokinetics”) of drug utilization. The epidemiological approaches to the study of the effects (or “pharmacodynamics”) of drug utilization, both beneficial and harmful, are covered in other chapters of this book.

TYPES OF DRUG UTILIZATION STUDY AND THEIR USES

Drug utilization studies may be *quantitative* or *qualitative*. In the former, the objective of the study is to quantify the present state, the developmental trends, and the time course of drug usage at various levels of the health care system, whether national, regional, local, or institutional. Routinely compiled drug statistics or drug utilization data that are the result of such studies can be used to estimate drug utilization in populations by age, sex, social class, morbidity, and other characteristics, and to identify areas of possible over- or underutilization. They also can be used as denominator data for calculating rates of reported adverse drug reactions; to monitor the utilization of specific therapeutic categories where particular problems can be anticipated (e.g., narcotic analgesics, hypnotics and sedatives, and other psychotropic drugs); to monitor the effects of informational and regulatory activities (e.g., adverse events alerts, delisting of drugs from therapeutic formularies); as markers for very crude estimates of disease prevalence (e.g., digitalis utilization for congestive heart failure, antiparkinsonian drugs for Parkinson’s disease); to plan for drug importation, production, and distribution; and to estimate drug expenditures.²

Qualitative studies, on the other hand, assess the appropriateness of drug utilization, usually by linking prescription data to the reasons for the

drug prescribing (see also Chapter 30). The crucial difference between these studies and quantitative drug utilization studies is that they include the concept of appropriateness.⁹ 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. Other possible criteria for poor drug prescribing include the failure to select a more effective or less hazardous drug if available, the use of a fixed combination drug when only one of its components is justified, or the use of a costly drug when a less costly equivalent drug is available.¹⁰ In North America, these studies are known as *drug utilization review* (DUR) or *drug utilization review studies*. For example, a large number of studies in North America have documented the extent of inappropriate prescribing of drugs, in particular antibiotics, and the associated adverse clinical, ecological, and economic consequences.^{11–24}

In Spain, the appropriateness of drug utilization has been assessed on the basis of adequate evidence for the clinical efficacy (“high intrinsic value”) of the most commonly sold drugs. The analysis revealed a striking proportion of drugs of “doubtful, no, or unacceptable value” among the 400 top pharmaceutical products in sales, albeit a trend toward more rational consumption as reflected in consumption of drugs of “high intrinsic value.”²⁵ More recently, this approach has been used to assess prescribing patterns in France, Germany, Great Britain, and Italy;²⁶ appropriateness of non-prescription drug sales in Brazil;²⁷ and drug prescribing in Spanish primary care centers.^{28,29} Another novel approach analyzed the number of drugs that accounted for 90% of drug use (DU90%) and the percentage of these drugs that adhered to the guideline issued by the Drug Committee in the catchment area.³⁰ The number of different products in the DU90% segment varied twofold among 24 primary health care centers in Stockholm; adherence to the guideline varied between 54 and 78%.

DUR and DUR studies are not interventions but rather activities aimed at problem detection and quantification. They should be distinguished,

Table 29.1. Drug utilization studies in perspective: operational concepts

	Drug statistics	Drug utilization study	Drug review
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

therefore, from DUR programs (Table 29.1) (see also Chapter 31). DUR studies are usually one-time projects, not routinely conducted. They provide for only minimal feedback to the involved prescribers and, most importantly, do not include any followup measures to ascertain whether any changes in drug therapy have occurred. A DUR program, on the other hand, is an intervention in the form of an authorized, structured, and *ongoing system* for improving the quality of drug use within a given health care institution. The quality of drug prescribing is evaluated by employing predetermined standards for initiating administrative or educational interventions to modify patterns of drug use which are not consistent with these standards. The measurement of the effectiveness of these interventions is an integral part of the program.^{9, 31, 32}

In the US, DUR programs (commonly known in hospitals as Drug Use Evaluation or DUE Programs) are part of the quality assurance activities required by Medicaid–Medicare regulations, the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), the former Professional Standards Review Organizations (PSRO), and Section 4401 of the Omnibus Budget Reconciliation Act of 1990^{9, 32, 33} (see Chapter 31). In Europe, DUR programs have taken the form of periodic drug utilization studies as elements of a “therapeutic audit” performed at various levels (patient, prescriber, hospital, county, municipality, country, and groups of countries), assessing not only the clinical consequences of drug utilization, but also the social and economic consequences. These are followed by whatever feedback is felt to be necessary and appropriate to effect changes in therapeutic practices.^{34–39} The therapeutic audit is performed on an epidemiological scale

and interventions (regulatory or educational) are aimed accordingly at whole populations or subgroups.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY RESEARCH

In order for a drug to be marketed, it must be shown that it can effectively modify the natural course of disease or alleviate symptoms when used appropriately, 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 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.⁴⁰

Studies in the United States,^{41, 42} Sweden,⁴³ and Denmark⁴⁴ have documented the importance of adverse drug reactions, as well as drug noncompliance, as causes of hospital admissions. Many of these drug related admissions are preventable, through the application of existing principles and data.⁴⁰ The situations that may lead to preventable adverse drug reactions and drug induced illness include the use of a drug for the wrong indication; the use of a potentially toxic drug when one with less risk of toxicity would be just as effective; the concurrent administration of an excessive number of drugs, thereby increasing the possibility of adverse drug interactions; the use of excessive doses, especially for pediatric or geriatric patients; and continued use of a drug after evidence

becomes available concerning important toxic effects. Many contributory causes have been proposed: excessive prescribing by the physician; failure to define therapeutic endpoints for drug use; the increased availability of potent prescription and nonprescription drugs; increased public exposure to drugs used or produced industrially that enter the environment; the availability of illicit preparations; and prescribers' lack of knowledge of the pharmacology and pharmacokinetics of the prescribed drugs.⁴⁰

Increased morbidity or mortality due to medication error,⁴⁵ poor patient compliance,⁴⁶ discontinuation of therapy,^{47,48} and problems in communication resulting from modern day fragmentation of patient care are also to be considered. The failure of physicians to prescribe an effective drug or effective doses for a treatable disease is a significant concern. For example, in a geographic area of Sweden with a higher suicide rate than average for the country, sales of antidepressant drugs were about half of those in other areas.⁴⁹ In the US, the underuse of β -blockers in elderly patients with myocardial infarction was associated with an increased risk of death.⁵⁰ Other studies have documented significant underuse of antithrombotic drugs,^{51–53} lipid-lowering therapy,^{54,55} beta-blockers,⁵⁶ aspirin,⁵⁷ and thrombolytics⁵⁸ in patients with appropriate indications, but the outcomes were not assessed.

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.^{59,60} Regardless, drug utilization studies address the relationship between therapeutic practice as recommended and actual clinical practice.⁶¹

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

A considerable amount of 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.⁶² 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, the data available are suboptimal—both the exposure data and the diagnosis data. In addition, the criteria to be used to define “appropriate” are arbitrary.

Since most statistics on drug consumption were compiled for administrative or commercial reasons, the data are usually expressed in terms of cost or volume (see Table 29.2). First, data on drug utilization can be available as total costs or unit costs, that is 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.⁶³

Volume data are also available, as the overall weight of the drug that is sold or the unit volume sold, that is the number of tablets, capsules, or doses sold. This is closer to the number of patients exposed. However, tablet sizes vary, making it difficult to translate weight into even the number of tablets. Prescription sizes also vary, so it is difficult to translate number of tablets into the number of exposed patients.

Table 29.2. Types of drug utilization datum available

- | |
|---|
| 1. Cost or unit cost |
| 2. Weight |
| 3. Number of tablets, capsules, doses, etc. |
| 4. Number of prescriptions |
| 5. Number of patients ingesting drug ^a |

^a Generally not available

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 further by the patients' actual drug intake, that is, their degree of compliance.

In the context of DUR, drug utilization data may be presented in the form of profiles of physicians according to the number, monetary value, and even type of prescription ordered during a given time period. Pharmacies may be ranked according to the number, cost, and type of prescription dispensed for similar intervals. However, these gross measures of prescription activity and drug use are very limited in their capacity to reflect the wide spectrum of specific problems in prescribing. For example, they ignore problems such as the wrong drug for the indication, the wrong drug for the patient, the wrong dose, the wrong interval, and the wrong duration of therapy. Also, one's deviation from the practices of the mean practitioner is not a good measure of one's "appropriateness" as a provider. Purely quantitative data characterizing prescribers as "high" or "low" may be driven, for example, by the number of patients seen by the physician and the type and severity of the patients' diseases. Likewise, cost profiles are not indicative of appropriateness, whether high or low relative to the mean.

From a qualitative perspective, to interpret drug utilization data appropriately, there is a need to relate the data to the reasons for the drug usage. Data on morbidity and mortality may be obtained from national registries (general or specialized); national samples where medical service reimbursement schemes operate; *ad hoc* surveys and special

studies; hospital records; physician records; and patient or household surveys. "Appropriateness" of use must be assessed relative to indication for treatment, patient characteristics (age, sex, habits), drug dosage (over- or underdosage), 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.^{64,65}

Generally agreed upon standards or criteria for appropriateness, based upon currently available knowledge, are essential elements of the drug utilization review process. These criteria must be based on scientifically established evidence; updated regularly according to new scientific evidence; explicitly stated (to ensure consistency in the evaluations); and applicable to a given setting.³¹ The development and standardization of these criteria are major undertakings. Finally, for drug utilization review programs, even the strategy to be used to optimize one's intervention is unclear.⁶⁶

CURRENTLY AVAILABLE SOLUTIONS

THE EVOLUTION OF DRUG UTILIZATION STUDIES

The current growth of interest in drug utilization studies began on both sides of the Atlantic in the early 1960s. Previously, drug utilization studies had been conducted mostly for marketing purposes and data were not widely available for use by academic researchers or health authorities. The increased interest resulted from recognition of the virtual explosion in the marketing of new drugs, the wide variations in the patterns of drug prescribing and consumption, the growing concern about the delayed adverse effects, and the increasing concern about the cost of drugs, as reflected in the increase in both the sales and the volume of prescriptions of drugs.^{35-38,63} However, the development of pharmacoepidemiological methods can be characterized by two different lines of work currently approaching each other from opposite

directions, strongly influenced by the varied availability and accessibility of data sources.

Drug utilization studies at the national and international levels have been more developed in Europe, where this line of research was pioneered by the Scandinavian countries, Scotland, and Northern Ireland. Under the auspices of the WHO Regional Office for Europe, a Drug Utilization Research Group was established in the 1970s to stimulate interest in comparative studies with a common methodology.^{35,67} Factors that contributed greatly to this line of development, primarily in the countries of Northern Europe, have been the relatively small size of the populations involved, the limited number of pharmaceutical products on the market (2000–3000 in Norway and Sweden), and the availability of centralized statistics on sales or prescriptions.³⁵ Drug utilization studies in Europe have been predominantly quantitative, describing and comparing patterns of utilization of specific groups of drugs according to geographic regions and time. For example, international studies have documented wide variations in the utilization of antidiabetic, psychotropic, and antihypertensive drugs among several European countries.^{35,36,68} Followup studies on the utilization of antidiabetic and antihypertensive drugs among some of these countries indicate that the differences cannot be explained only by differences in the prevalence of disease.^{69–71} National studies have also revealed striking variations in drug utilization among regions and communities within the same country.³⁵ More recently, one study addressed the relation between variations in drug sales and treatment outcomes. For example, the degree of good metabolic control, as defined by the authors, in diabetic subjects in three Swedish areas with high, medium, and low sales of antidiabetic drugs was achieved among only 16, 17, and 12% of subjects, respectively.⁷²

In the US, drug utilization research has developed on a smaller scale, primarily at institutional or local health program levels. Factors that have hindered studies at a national level have been the size of the population, the number of pharmaceutical products on the market (20 000 to 30 000), and the lack of an all-encompassing pharmaceutical data collection system.⁷³ Data on drug use are more readily available from prepaid health plans,

health delivery institutions, and public health care programs. For example, early studies of physician prescribing showed that prescribing patterns varied greatly among physicians, according to their place and type of practice and the community in which they prescribed.⁷⁴ In US drug utilization research, greater emphasis has been placed on the study of the quality of physician prescribing habits, in particular with respect to antibiotics, in both hospital and outpatient settings.^{11–24} However, studies of the national patterns of drug utilization and expenditures in the US have also been published.^{75,76}

Because of the critical importance of the decision making process in drug prescribing, a number of studies have addressed the factors that influence this decision: education, advertising, colleagues, working circumstances, personality, control and regulatory measures, and demands from society and patients.^{77,78} Some controversy exists concerning the relative impact of the various sources of influence on prescribing behavior, particularly the influence of pharmaceutical advertising. In studies of hospital practice the following factors have been stated to contribute to excessive or inappropriate prescribing: simple errors of omission; physician ignorance of cost issues in prescribing; failure to review medication orders frequently and critically; inability to keep up to date with developments in pharmacology and therapeutics; insulation of physicians and patients from cost considerations because of third party coverage; and lack of communication between physicians and pharmacists.⁷⁹

The intervention strategies aimed at improving prescribing behavior in hospital as well as primary care settings have been critically reviewed.^{79–81} These include (discussed in Chapter 30) dissemination of printed educational materials alone, multimedia warning campaigns, drug utilization audit followed by mailed or interactive feedback of aggregated results, group education through lectures or rounds, use of computerized reminder systems, use of opinion leaders to informally “endorse” or support specific behavior change interventions, one-to-one education initiated by a drug utilization expert, and required consultation or justification prior to the use of specific drugs.

CURRENT DATA SOURCES

Currently available computer databases for studies of drug utilization may be classified as non-diagnosis-linked and diagnosis-linked (see Table 29.3). Most of these data sources lack information on morbidity and are mostly used for generating drug statistics and descriptive studies of patterns of drug consumption. Some collect data in the form of drug sales (e.g., the National Corporation of Pharmacies in Sweden, Norwegian Drug Monopoly, and the National Agency for Medicines and Social Insurance in Finland),^{82–84} drug movement at various levels of the drug distribution channel (IMS America's National Prescription Audit, US Pharmaceutical Market—Hospitals, US Pharmaceutical Market—Drugstores),⁷³ pharmaceutical or medical billing data (Prescription Pricing Authority in the UK, Spain's Drug Data Bank, Medicaid Management Information System),^{35, 85, 86} or samples of prescriptions (Swedish Prescription Survey).⁸⁷

The County of Jämtland Project (Sweden) is of interest for longitudinal patient-specific studies of drug utilization.^{49, 88, 89} All drug prescriptions dispensed to 13% of the Jämtland population

(approximately 17 000) have been continuously monitored since 1970. The recorded information includes the patient's unique identity number; name, dosage, quantity, and price of the drug; date of dispensing; dispensing pharmacy; and prescribing physician. Information relating to morbidity (diagnoses), however, is missing. Unfortunately, because of sensitivity to the issue of data confidentiality in Sweden, the correspondingly recorded data relative to individual patients in other parts of Sweden is not available for use in health care audits.^{64, 89}

The Odense Pharmacoepidemiologic Database (OPED) and the Pharmacoepidemiologic Prescription Database of the County of North Jutland are two similar databases that include about half a million inhabitants in Denmark.^{90, 91} These databases contain all dispensed prescriptions since the early 1990s. The following information is captured for each prescription: a unique person identifier, the date of dispensing, identification of the dispensed product, the pharmacy, and the prescriber. The databases do not include information on over-the-counter medications (laxatives, analgesics, ibuprofen, antihistamines, antitussives, and certain anti-ulcer drugs) or nonsubsidized

Table 29.3. Some computer databases for drug utilization studies

Not diagnosis-linked	Diagnosis-linked
North America	
National Prescription Audit ^a	National Disease and Therapeutic Index ^a
US Pharmaceutical Market—Drugstores ^a	Kaiser Permanente Medical Plan ^b
US Pharmaceutical Market—Hospitals ^a	Group Health Cooperative of Puget Sound ^b
Medicaid Management Information Systems	COMPASS ^{®b, c}
Saskatchewan Health Plan ^b	DURbase ^{®b, c}
Europe	
Swedish National Corporation of Survey Pharmacies Project ^b	Sweden's Diagnosis and Therapy
Sweden's Prescription Survey	Sweden's Community of Tierp
Sweden's County of Jämtland Project	
Norwegian Medicinal Depot	
United Kingdom's Prescription Pricing Authority	
Spain's Drug Data Bank (National Institute of Health)	
Denmark's Odense Pharmacoepidemiologic Database	
Denmark's Pharmacoepidemiologic Prescription Database of the County of North Jutland	

^a IMS America, Ltd.

^b Patient-specific data available for longitudinal studies.

^c Health Information Designs, Inc.

drugs (oral contraceptives, hypnotics, and sedatives). They have been used for a number of population based pharmacoepidemiologic surveys such as the use of the new antidepressants, inappropriate use of inhaled steroids in asthma treatment, inappropriate use of sumatriptan, hemorrhagic complication during oral anticoagulant therapy and low use of long term hormone replacement therapy.⁹²⁻⁹⁶ The OPED database has also been used to develop a graphical approach to reduce the overwhelming volume of data in population based pharmacoepidemiological databases into a few parameters and a "waiting time distribution" that can be used to screen for certain unusual or unexpected patterns of drug use.⁹⁰ Based on prescriptions dispensed to individual patients, key parameters such as incidence, one-year and point prevalence, duration of treatment, relapse rate, and seasonality receive a visual correlate.

Several databases contain both drug and morbidity data but have been used to a relatively limited extent for this type of study, as opposed to hypothesis testing studies. These include data from the Group Health Cooperative of Puget Sound and the Kaiser Permanente Medical Care Programs, described in more detail in Chapters 15 and 16, respectively.

The National Disease and Therapeutic Index (NDTI), by IMS America, Ltd., is an ongoing study of physician prescribing conducted mainly for use by pharmaceutical companies in their marketing activities. This study employs a rotating sample of office-based physicians who record all patient encounters and corresponding "drug mentions" for two-day periods four times a year. A special prescription form is used to collect information on the drug (specific product, dosage form, new versus continuing therapy), patient characteristics (sex), prescriber (specialty, location, region), type of consultation (first versus subsequent), concomitant drugs and diagnoses, and the desired pharmacological action.⁷⁶ Data have been made available to academic researchers and the US Food and Drug Administration.^{75, 76, 97, 98} Although useful for studies of prescribing, longitudinal patient-specific studies are not possible with this database.

Similar to IMS America's NDTI, the Swedish Diagnosis and Therapy Survey is a collaborative project run by Swedish Pharmaceutical Data Ltd (LSAB), the National Corporation of Pharmacies (Apoteket), the Swedish Medical Association, and the National Board of Health and Welfare. A random sample of one out of eight physicians (2200 per year) is asked to participate (non-participation rate estimated at 25%). During a study week all prescriptions are recorded on self-copying forms with additional information on indication/diagnosis for each drug, patient age, sex, type of consultation, etc. The data are analyzed according to the IMS routines and are available as printouts to subscribers twice a year.⁹⁹ Drug utilization data from this ongoing survey, in combination with overall sales statistics, are published yearly by the National Corporation of Pharmacies⁸² and also made available for research.¹⁰⁰

The Community of Tierp Project is run by the Center for Primary Care Research, University of Uppsala, Sweden. Prescription and morbidity data have been routinely collected from all pharmacies and the health center within the community for all residents since 1972.¹⁰¹ Although diagnosis information is collected in this project, studies published in English have not exploited this aspect of these data. They have been limited to using the database to identify and study drug utilization patterns in patients with certain characteristics, such as "heavy use," or users of antidiabetics or psychotropic drugs.^{102, 103} Limitations of this database are the small size of the population covered (21 000 persons) and questions regarding the representativeness of this community for the whole of Sweden. Moreover, aggregate drug data are kept on a pharmacological or therapeutic class basis so, with the exception of the benzodiazepines, it is difficult to study individual drugs.

In Canada, the province of Saskatchewan has a series of computerized databases describing health services paid for by the provincial Department of Health, including prescription drugs.¹⁰⁴ A variety of drug utilization studies have been performed using these data, which are described in more detail in Chapter 20.

In the United States DURbase[®] and COMPASS[®] are two databases developed by Health Information Designs, Inc. Both are based on Medicaid medical care and pharmaceutical billing data from 11 out of the 50 US states.^{105,106} Selected portions of these data are processed to compile patient-specific diagnosis and drug profiles containing the following information: patient age, sex, race, county of residence, drug information (identity, dosage form, strength, quantity dispensed), prescriber code, dispensing pharmacist code, diagnosis, and corresponding dates of services provided. DURbase[®] is the drug utilization review program for some of these state Medicaid programs, and for other non-Medicaid programs. Other commercial firms offer similar services. COMPASS[®] uses the data obtained from DURbase[®] for pharmacoepidemiology studies (see Chapter 19). Drug utilization studies performed using COMPASS[®] have been limited. With the skewed population included in COMPASS[®], the generalizability of the results would be of concern.

Medical and pharmaceutical computer databases are generally not available outside North America and Europe. An indicator-based approach, developed in the early 1990s by the International Network for Rational Use of Drugs (INRUD) and WHO,¹⁰⁷ has facilitated the study of drug utilization in developing countries. It includes recommendations on minimum sample sizes, sampling methods, and data collection techniques, depending on study objectives. The methodology recommends 12 core indicators and seven complementary indicators to study drug use in health facilities (Table 29.4). These indicators can be used to describe prescribing practice,¹⁰⁸ conduct monitoring and supervision,¹⁰⁹ and assess the impact of interventions.^{110–112} To date, researchers in more than 25 countries of Africa, Asia, and Latin America have used this methodology.

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

Table 29.4. WHO/INRUD drug use indicators

Core indicators
Prescribing indicators
Average number of drugs per encounter
Percentage of drugs prescribed by generic name
Percentage of encounters with an antibiotic prescribed
Percentage of encounters with an injection prescribed
Percentage of drugs prescribed from essential drugs list or formulary
Patient care indicators
Average consultation time
Average dispensing time
Percentage of drugs actually dispensed
Percentage of drugs adequately labeled
Patient's knowledge of correct dosage
Facility indicators
Availability of copy of essential drugs list or formulary
Availability of key drugs
Complementary indicators
Percentage of patients treated without drugs
Average drug cost per encounter
Percentage of drug costs spent on antibiotics
Percentage of drug costs spent on injections
Prescription in accordance with treatment guidelines
Percentage of patients satisfied with care they received
Percentage of health facilities with access to impartial drug information

sales statistics or pharmacy inventory data (quantity of packages, tablets, or other dosage forms) into medically meaningful units, to make crude estimates of the number of persons exposed to a particular drug or class of drugs.^{35,63} The DDD is the assumed average maintenance dose for the main indication of a particular drug. Expressed as DDDs per 1000 inhabitants per day, it can be interpreted as the proportion of the population that may receive treatment with a particular drug.^{113–115} For use in hospital settings, it may be expressed as DDDs per 100 bed days (adjusted for occupancy rate).^{116–118} The method has been useful in describing and comparing patterns of drug utilization,^{35,36} providing denominator data to estimate adverse drug reaction rates,¹¹⁹ performing epidemiological screening for problems in drug utilization,³⁹ and monitoring the effects of informational and regulatory activities.^{2, 120, 121}

The advantages of this methodology include: its usefulness for working with readily available gross drug statistics at various levels of the health chain;

as a standardized unit of measurement, it allows comparisons between drugs in the same therapeutic class and between different health care settings or geographic areas, and evaluations of trends over time; and it is relatively easy and inexpensive to use. The methodology is being used by an increasing number of researchers worldwide.^{35, 36, 117, 118, 121–129}

The DDD methodology has, however, some significant limitations. The DDD is not a recommended dose, but rather a technical unit of comparison. Since the standard DDDs are based on usage in the Nordic countries, some DDDs may be high or low relative to other countries. Many drugs that are not marketed in the Nordic countries have not been assigned DDDs, although guidelines have been published for defining DDDs under these circumstances.¹¹⁵ Additional problems arise when dosages vary widely (antibiotics), when one drug is used for more than one major indication (aspirin), or when drugs are used in combination with other drugs for the same disease. Moreover, the DDD does not take into account pediatric uses. Since children's doses are substantially lower than the established DDDs, this situation will lead to an underestimation of population exposures, which may be significant in countries with a large pediatric population. However, pediatric DDDs have also been proposed.¹³⁰ 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.¹³¹ Problems may arise in calculating the PDD due to a lack of clear and exact dosage indication in the prescription, as is often the case with the prescribing of insulin. Prescriptions for chronic therapy, as in the case of insulin, may be refilled many times and the dosage may be altered verbally between prescribing events.¹³² For certain groups of drugs, such as the oral antidiabetics, the mean PDD may be lower than the corresponding DDDs. Up to twofold variations in the mean PDD have been documented in international comparisons.¹³¹ Higher PDDs have been observed in the US relative to Sweden for commonly prescribed

drugs, such as hydrochlorothiazide, diazepam, and oxazepam.^{133–135} In risk assessments of antidepressants among suicides, a refined person year of use estimate was obtained from adjusting the DDD by the average PDD for individual antidepressants.^{136, 137} Although the DDD and the PDD may be used to estimate population drug exposure, the methodology is not useful to quantify or identify patients who receive doses lower or higher than those considered effective and safe.

CLASSIFICATION SYSTEMS

The Anatomic Therapeutic Chemical (ATC) classification system is generally used in conjunction with the DDD methodology.¹¹⁵ The Norwegian Medicinal Depot, currently a WHO Collaborating Centre for Drug Statistics Methodology, developed it. The ATC system is based on the main principles of the Anatomical Classification system developed by the European Pharmaceutical Market Research Association (EPHRA) and the International Pharmaceutical Market Research Group (IPMRG).

The ATC system consists of five hierarchical levels: a main anatomical group, two therapeutic subgroups, a chemical–therapeutic subgroup and a chemical substance subgroup. The coding of furosemide preparations is used to illustrate the ATC classification structure in Table 29.5. The first three levels are modifications of the three-level EPHRA and IPMRG classification system. The fourth and fifth levels are extensions developed by the Norwegian Medicinal Depot. On-going discussions aim to identify differences in the two classification systems and harmonize the first three levels. Statistics reported with the ATC system should not be directly compared with figures prepared with the EPHRA system.

The ATC system has a number of limitations. Drugs that have two or more equally important indications can only use one ATC code. For example, there is no ATC code for propranolol use as an antiarrhythmic drug; the only ATC codes available are as an antihypertensive or antimigraine drug. Propranolol may also be coded differently in one country if its use is mainly as an antihypertensive while in another it is used

Table 29.5. ATC and IDIS classification and coding structures for furosemide

ATC Classification (C03CA01)	
C	Cardiovascular system (first level, main anatomical group)
03	Diuretics (second level, main therapeutic group)
C	High ceiling diuretics (third level, therapeutic subgroup)
A	Sulfonamides, plain (fourth level, chemical therapeutic subgroup)
01	Furosemide (fifth level, chemical substance)
IDIS classification (40280401)	
40	Electrolyte solutions (first level, main therapeutic group)
28	Diuretics (second level, therapeutic subcategory)
04	Loop diuretics (third level, therapeutic subcategory)
01	Furosemide (fourth level, chemical substance)

mainly as an antimigraine drug. Some drugs may be coded differently if the therapeutic use is associated with the dosage form or strength. For example, prednisolone may use one code for dermatological products and another for systemic use products; one strength of clonidine may be classified as an antihypertensive and another as an antimigraine drug; and one strength of medroxyprogesterone may be classified as endocrine therapy for cancer treatment while another strength is coded as a sex hormone. Fixed dose combination products may also pose difficulties. For example, a combination product that contains an analgesic and a tranquilizer is classified as an analgesic, even though it also contains a psychotropic substance.

The European Drug Utilization Research Group (EURO DURG), formerly WHO Drug Utilization Research Group and currently an association of national Drug Utilization Research Groups, recommends the use of the ATC classification system for reporting drug consumption statistics and conducting comparative drug utilization research. This has been implemented in the five Nordic countries, Australia, and some European countries.^{114, 128, 129} The WHO International

Drug Monitoring Program uses the system for drug coding in adverse drug reaction monitoring. Some developing countries have begun to use the ATC system to classify their essential drugs; this may eventually lead to preparation of drug utilization statistics.¹³⁸⁻¹⁴¹

In the US, the Iowa Drug Information System (IDIS) is a hierarchical drug coding system that is based on the three therapeutic categories of American Hospital Formulary Society (AHFS), to which a fourth level was added to code individual drug ingredients.¹⁴² The IDIS code has eight numeric digits, two digits per level (see Table 29.5). This coding system was used in the Established Populations for Epidemiologic Studies of the Elderly survey. Other coding systems such as the National Drug Code and the Veterans' Administration Classification¹⁴³ do not provide unique codes for drug ingredients.

INTERVENTION STRATEGIES BASED ON DRUG UTILIZATION DATA

Numerous studies have described interventions aimed at improving prescribing by the use of drug utilization data obtained from qualitative drug utilization studies, and are discussed more in Chapter 30. Two innovative intervention strategies exemplify different approaches to the use of drug utilization data available from computer databases of office practice.

In a randomized clinical trial, Avorn and Soumerai¹⁴⁴ used data from the Medicaid Management Information System to identify physicians who were prescribing drugs that were assessed as inappropriate (based on considerations of documented efficacy, relative efficacy, and relative cost). These physicians were targeted for educational or information activities, as either face-to-face contacts or written drug information. Schaffner *et al.*¹⁴⁵ and Ray *et al.*¹⁴⁶ used a similar approach in another controlled intervention study comparing different strategies aimed at modifying physician prescribing behavior: written drug information versus personal visits by pharmacists versus personal visits by physician educators. These two studies demonstrated the efficacy of face-to-face methods in improving drug prescribing.

The second approach uses claims data to perform computerized screening for patients who may be at increased risk for drug-induced illness, using patient-specific medical and drug histories.^{106, 147, 148} Health professionals then evaluate profiles of patients with possibly inappropriate drug use. If considered appropriate, a letter is sent to the prescriber providing a profile of the patient's relevant computerized claims record and a warning of the potential for drug-induced disease. Often the problem is a concomitant drug or diagnosis that the prescriber was unaware of. This approach is obviously much less expensive than the face-to-face approach. Using before and after comparisons, a significant reduction in drug-induced hospitalizations has been noted.¹⁴⁷ However, the interpretation of these results is hampered by the use of a non-experimental design. A simultaneously controlled trial is needed to adequately assess the value of this approach. (See Chapter 31 for more information about computerized claims-based drug utilization review programs.)

Many other studies have described intervention strategies based on providing drug utilization data feedback, alone or in combination with printed material and/or other "educational strategies," for example group discussions, lectures, seminars, or personal visits by "experts." The results from these studies are conflicting. Some suggest that methods that involve only feedback of drug utilization data or audit results are ineffective. Others suggest a transient effectiveness for those that combine the use of drug utilization review data with group discussions, lectures, and visits by "experts." However, these are difficult to interpret because of limitations in their research designs.⁷⁹

Conceptually, DUR programs are aimed at the improvement of medical care and cost containment. However, in practice traditional approaches have focused on the control of abuse or overuse of drugs, polypharmacy, or patients obtaining prescriptions from many different prescribers. Additionally, most DUR studies have emphasized process measures of quality of care, for example the use of clinical laboratory tests to monitor for adverse effects during chloramphenicol or aminoglycoside therapy. The approach described by Strom *et al.*,¹⁰⁶ Morse *et al.*,¹⁴⁷ and Groves¹⁴⁸ is a

significant advance in DUR programs, as it is primarily aimed at improving measurable patient outcomes. Also, it does not impose arbitrary restrictions on drug use, potentially impairing patient care, but seeks to reduce costs by improving patient care. In seeking to reduce the financial impact of drug use, it does not focus on the drug costs themselves, but on the effects of the drugs. By reducing the need for medical care through the beneficial effects of drugs, or by increasing the need for remedial medical care because of drug toxicity, pharmaceuticals can have a financial impact on the health care system that is much larger than the cost of the drugs themselves. (This is discussed more in Chapter 35.) Despite their appeal, however, the true impact of such programs on the quality of care still remains to be established. (See Chapter 31 for a detailed discussion of DUR.)

THE FUTURE

OPPORTUNITIES

From a public health perspective, the observed differences in national and international patterns of drug utilization require much further study. Among other things, the medical consequences as well as the explanations for such differences are still not well documented.

Numerous studies have addressed the factors influencing drug prescribing. However, the relative importance of the many determinants of appropriate prescribing still remains to be adequately elucidated. Further research is needed to better define to what degree and which determinants of inappropriate prescribing are susceptible to modification and what might be an appropriate mix of interventions to achieve optimal impact. Although regulation is effective, it is not possible to regulate all aspects of the clinical decision making process to ensure optimal drug prescribing.¹⁴⁹ Other approaches in addition to educational and informational measures need to be explored.

Many strategies aimed at modifying prescribing behavior have been proposed and adopted. The evidence to date indicates that mailed educational

materials alone are not sufficient to modify prescribing behavior.⁸⁰ In addition, a recent Australian study concluded that mailed, unsolicited, centralized, government-sponsored feedback based on aggregate prescribing data had no impact on the prescribing levels of general practitioners.¹⁵⁰ For interventions that have been shown to be effective in improving drug prescribing (discussed in Chapter 30), there is a need to further define their relative efficacy and proper role in a comprehensive strategy for optimizing drug utilization. Questions yet to be addressed through proper methodology deal with the role of printed drug information such as drug bulletins, the duration of effect of educational interventions such as group discussions, lectures, and seminars, each in both the outpatient as well as the inpatient settings, and the generalizability of face-to-face methods as described by Avorn and Soumerai,¹⁴⁴ Schaffner *et al.*,¹⁴⁵ and Ray *et al.*¹⁴⁶

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 (see Chapter 31). Although numerous studies have described the results of these and other novel programs, documentation of their efficacy in improving quality of care is an important subject for future work. Patient outcome measures as well as process measures of quality of drug utilization have to be included in such studies. To be effective and efficient, health care policy options should be based on sound scientific evidence.¹⁵¹

PROBLEMS

The use of computerized databases has greatly facilitated the study of drug utilization. Although useful, most of these databases are far from ideal, as they have been set up mainly for administrative purposes, such as reimbursement, and drug utilization data are obtained as “spin off” information. The model information system that will suit both medical and administrative needs¹⁵² is not to be expected in the near future. Existing medical and pharmaceutical databases, with all

their described limitations, will continue to be the main resources for these drug utilization studies.

Confidentiality of patient records has been successfully handled at the technical level. However, in many countries political acceptance may be much more difficult to achieve. For example, although patient-specific information is captured in some of the current databases in Sweden, due to legal restrictions this valuable information is not saved or stored and, thus, not available for health services research. EURO DURG researchers have reported difficulties arising from confidentiality laws in five of 10 European countries.¹⁵³ It is feared that implementation of Directive 95/46/EC of 24 October 1995, on the protection of individuals with regard to the processing of personal data and on the free movement of such data in the European Union, may adversely affect researcher access to patient health data (see also Chapter 26).

In an era of increased interest in cost containment and cost-effectiveness, research may not be awarded high priority, resulting in reduced opportunities for financing much needed drug utilization research. Moreover, the recruitment and training of researchers for this relatively new field may be hampered by limitations in funding, as well as limitations in career opportunities. These two problems will impose significant constraints on the future development of studies in drug utilization. However, despite this, the search must continue for simple and relatively inexpensive methods for conducting descriptive studies of drug utilization and effective intervention strategies that may contribute to the optimization of drug therapy. Fortunately, the increasing commitment to drug utilization research is reflected in the development and growth of international groups such as the International Society for Pharmacoepidemiology (ISPE),¹⁵⁴ the International Clinical Epidemiology Network (INCLIN),¹⁵⁵ the European Drug Utilization Research Group (EURO DURG),¹⁵⁶ and the International Network for Rational Use of Drugs (INRUD).^{157, 158}

In summary, the study of drug utilization is still evolving. Studies in Europe may be characterized as large scale—regional, national, and international—but primarily quantitative studies of descriptive epidemiology, exploiting relatively

cheap and readily available sources of drug statistics. In North America, most studies have been of more limited geographic scope, but have addressed the issues of appropriateness of drug prescribing, based primarily on process measures of quality of care. 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 the expansion of this field of study. The WHO/INRUD indicator-based approach to drug utilization studies is facilitating the development of drug utilization research in developing countries. Many strategies have already been proposed and are being implemented to improve the quality of drug prescribing. Drug utilization review programs, particularly approaches that take into primary consideration patient outcome measures, are one such strategy that merits 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, may slow the growth of drug utilization research.

REFERENCES

1. WHO Expert Committee. *The Selection of Essential Drugs (Technical Report Series 615)*. Geneva: World Health Organization, 1977.
2. Lunde PKM, Andrew M, Baksaas I. Drug utilization—an instrument in drug research. In: Kewitz H, Roots I, Voight K, eds. *Epidemiological Concepts in Clinical Pharmacology*. Berlin: Springer 1987; 57.
3. Brodie DC. *Drug Utilization and Drug Utilization Review and Control*, NCHS-RD-70-8. Rockville: Department of Health, Education, and Welfare, National Center for Health Services Research and Development, Health Services and Mental Health Administration, 1970.
4. Conley BE. Social and economic aspects of drug utilization research. Hamilton, Drug Intelligence, 1976.
5. Baksaas I, Lunde PKM. Drug utilization: pharmacokinetics in the community. *Trends Pharmacol Sci* 1981; **2**: 5–7.
6. Tognoni G. Drug use and monitoring. In: Holland WW, ed. *Evaluation of health care*. London: Oxford University Press, 1983; 207.
7. Laporte JR, Tognoni G, eds. *Principios de epidemiologia del medicamento*, 2nd edn. Barcelona: Ediciones Científicas y Técnicas, 1993.
8. Lunde PKM, Baksaas I. Epidemiology of drug utilization—basic concepts and methodology. *Acta Med Scand Suppl* 1988; **721**: 7–11.
9. Stolar MH. Drug use review: Operational definitions. *Am J Hosp Pharm* 1978; **35**: 76–8.
10. Lee PR. America is an overmedicated society. In: Lasagna L, ed. *Controversies in therapeutics*. Philadelphia, PA: Saunders, 1980, 4–15.
11. Scheckler NE, Bennet JV. Antibiotic usage in seven community hospitals. *J Am Med Assoc* 1970; **213**: 264–7.
12. Roberts WA, Visconti AJ. The rational and irrational use of systemic antimicrobial drugs. *Am J Hosp Pharm* 1972; **29**: 828–34.
13. Gibbs CW Jr, Gibson JT, Newton DS. Drug utilization review of actual versus preferred pediatric antibiotic therapy. *Am J Hosp Pharm* 1973; **30**: 892–7.
14. Pierpaoli PG, Coarse JF, Tilton RC. Antibiotic use control—an institutional model. *Drug Intell Clin Pharm* 1976; **10**: 258–67.
15. Castle M, Wilfert CM, Cate TR, Osterhout S. Antibiotic use at Duke University Medical Center. *J Am Med Assoc* 1977; **237**: 2819–22.
16. Perry TL, Guyatt GH. Antimicrobial drug use in three Canadian hospitals. *Can Med Assoc J* 1977; **116**: 253–6.
17. Counts G. Review and control of antimicrobial usage in hospitalized patients. *J Am Med Assoc* 1977; **238**: 2170–2.
18. Achong MR, Hauser BA, Krusky JL. Rational and irrational use of antibiotics in a Canadian teaching hospital. *Can Med Assoc J* 1977; **116**: 256–9.
19. Maki D, Schuna AA. A study of antimicrobial misuse in a university hospital. *Am J Med Sci* 1978; **275**: 271–82.
20. Shapiro M, Townsend TR, Rosner B, Kass EH. Use of antimicrobial drugs in general hospitals. *New Engl J Med* 1979; **301**: 351–3.
21. Shapiro M, Townsend TR, Rosner B, Kass EH. Use of antimicrobial drugs in general hospitals. II. Analysis of patterns of use. *J Infect Dis* 1979; **139**: 698–706.
22. Townsend TR, Shapiro M, Rosner B, Kass EH. Use of antimicrobial drugs in general hospitals. IV. Infants and children. *Pediatrics* 1979; **64**: 573–8.
23. Schollenberg E, Albritton WL. Antibiotic misuse in a pediatric teaching hospital. *Can Med Assoc J* 1980; **122**: 49–52.
24. Naqvi SH, Dunkle LM, Timmerman KJ, Reichley RM, Stanley DL, O'Connor D. Antibiotic usage in

- a pediatric medical center. *J Am Med Assoc* 1980; **242**: 1981–4.
25. Laporte JR, Porta M, Capellá D. Drug utilization studies: A tool for determining the effectiveness of drug use. *Br J Clin Pharmacol* 1983; **16**: 301–4.
 26. Garattini S, Garattini L. Pharmaceutical prescriptions in four European countries. *Lancet* 1993; **342**: 1191–2.
 27. Heineck I, Schenkel EP, Vidal X. Medicamentos de venta libre en el Brasil. *Rev Panam Salud Pública* 1998; **3**: 385–91.
 28. Casado Zuriguel I. Intervención para mejorar la calidad de prescripción de antibióticos en una area básica de salud. *Aten Primaria* 1993; **11**: 37–9.
 29. Mata Cases M, Avellana Revuelta E, Davins Miralles J, Calvet Junoy S, Hortelano Garca MA, Martn López A. Mejora continua de la calidad de la prescripción crónica en un centro de atención primaria: seguimiento de 5 años. *Aten Primaria* 1994; **13**: 172–6.
 30. Bergman U, Popa C, Tomson Y, Wettermark B, Einaron TR, Åberg H, Sjöqvist F. Drug utilization 90%—a simple method for assessing the quality of drug prescribing. *Eur J Clin Pharmacol* 1998; **54**: 113–8.
 31. Rucker TD. Drug utilization review: Guidelines for program development. In: Allozo JL, ed. *Clinical and social pharmacology*. Postmarketing period. Aulendorf: Cantor, 1985; 57.
 32. Todd M. Drug use evaluation. In: Brown TR, ed. *Handbook of institutional pharmacy practice*, 3rd edn. Bethesda, MD: American Society of Hospital Pharmacists, 1992; 261–71.
 33. Schulke DG. A congressional perspective on inappropriate drug therapy and drug utilization review. *Clin Pharmacol Ther* 1991; **50**: 606–11.
 34. Crooks J. Methods of “audit” in drug use. In: Duchêne-Marullaz P, ed. *Advances in pharmacology and therapeutics: Vol. 6. Clinical pharmacology*. Oxford, Pergamon, 1978; 189–95.
 35. Bergman U, Grímsson A, Wahba AHW, Westerholm B, eds. *Studies in drug utilization (European series 8)*. Copenhagen: WHO Regional Office for Europe, 1979.
 36. Sjöqvist F, Agenäs I, eds. *Drug utilization studies: Implications for medical care*. *Acta Med Scand* 1984; suppl 683.
 37. Baksaas I, Lunde PKM. National drug policies: The need for drug utilization studies. *Trends Pharmacol Sci* 1986; **7**: 331–4.
 38. Sjöqvist F, Boëthius G. Attitudes to development of drug therapy in Scandinavia. *Br J Clin Pharmacol* 1986; **22**: S19–26.
 39. Westerholm B. Therapeutic auditing at the national and international level. *Br J Clin Pharmacol* 1986; **22**: 55S–9S.
 40. Melmon KL. Preventable drug reactions: Causes and cures. *New Engl J Med* 1971; **284**: 1361–3.
 41. Morse ML, LeRoy A. *Impact of inappropriate ambulatory patient drug therapy on hospitalizations*. Washington, DC: Department of Health, Education, and Welfare, Health Care Financing Administration, 1978.
 42. Steel KL, Gertman PM, Crescenzi C, Anderson J. Iatrogenic illness on a general medical service at a university hospital. *New Engl J Med* 1981; **304**: 638–40.
 43. Bergman U, Wiholm BE. Drug related problems causing admission to a medical clinic. *Eur J Clin Pharmacol* 1981; **20**: 193–200.
 44. Davidsen F, Haghfelt T, Gram LF, Brøsen K. Adverse drug reactions and drug non-compliance as primary causes of admission to a cardiology department. *Eur J Clin Pharmacol* 1988; **34**: 83–6.
 45. Bates DW. Medication errors: how common are they and what can be done to prevent them? *Drug Safety* 1996; **15**: 303–10.
 46. Haynes RB, Taylor DW, Sackett DL, eds. *Compliance in health care*. Baltimore, MD: The Johns Hopkins University Press, 1979.
 47. Andrade SE, Walker AM, Gottlieb LK, et al. Discontinuation of antihyperlipidemic drugs—do rates reported in clinical trials reflect rates in primary care settings? *New Engl J Med* 1995; **332**: 1125–31.
 48. Avorn J, Monette J, Lacour A, et al. Persistence of use of lipid-lowering medications: a cross-national study. *J Am Med Assoc* 1998; **279**: 1458–62.
 49. Isacson G, Boëthius G, Bergman U. Low level of antidepressant prescription for people who later commit suicide: 15 years of experience from a population-based drug database in Sweden. *Acta Psychiatr Scand* 1992; **85**: 444–8.
 50. Soumerai SB, McLaughlin TJ, Spiegelman D, et al. Adverse outcomes of underuse of beta-blockers in elderly survivors of acute myocardial infarction. *J Am Med Assoc* 1997; **277**: 115–21.
 51. Whittle J, Wickenheiser L, Venditti LN. Is warfarin underused in the treatment of elderly persons with atrial fibrillation? *Arch Intern Med* 1997; **157**: 441–5.
 52. Antani MR, Beyth RJ, Covinsky KE, et al. Failure to prescribe warfarin to patients with nonrheumatic atrial fibrillation. *J Gen Intern Med* 1996; **11**: 713–20.
 53. Albers GW, Yim JM, Belew KM, Bittar N, Hattemer CR, Phillips BG, Kemp S, Hall EA, Morton DJ, Vlasses PH. Status of antithrombotic therapy for patients with atrial fibrillation in university hospitals. *Arch Intern Med* 1996; **156**: 2311–6.
 54. Mendelson G, Aronow WS. Underutilization of measurement of serum low-density lipoprotein

- cholesterol levels and of lipid-lowering therapy in older patients with manifest atherosclerotic disease. *J Am Geriatr Soc* 1998; **46**: 1128–31.
55. Aronow S. Underutilization of lipid-lowering drugs in older persons with prior myocardial infarction and a serum low-density lipoprotein cholesterol >125 mg/dl. *Am J Cardiol* 1998; **82**: 668–9, A6, A8.
 56. McLaughlin TJ, Soumerai SB, Willison DJ, *et al*. Adherence to national guidelines for drug treatment of suspected acute myocardial infarction: evidence for undertreatment in women and the elderly. *Arch Intern Med* 1996; **156**: 799–805.
 57. Krumholtz HM, Radford MJ, Ellerbeck EF, *et al*. Aspirin in the treatment of acute myocardial infarction in elderly Medicare beneficiaries. Patterns of use and outcomes. *Circulation* 1995; **92**: 2841–7.
 58. O'Connor GT, Quinton HB, Traven ND, *et al*. Geographic variation in the treatment of acute myocardial infarction: the Cooperative Cardiovascular Project. *J Am Med Assoc* 1999; **281**: 627–33.
 59. Slone D, Shapiro S, Miettinen OS, *et al*. Drug evaluation after marketing. *Ann Intern Med* 1979; **90**: 257–61.
 60. Strom BL, Melmon KL, Miettinen OS. Postmarketing studies of drug efficacy. *Arch Intern Med* 1985; **145**: 1791–4.
 61. Lasagna L. A plea for the naturalistic study of medicines. *Eur J Clin Pharmacol* 1974; **7**: 153–4.
 62. Rucker TD. Data, sources, and limitations. *J Am Med Assoc* 1974; **230**: 888–90.
 63. Anonymous. *Consumption of Drugs: Report on a Symposium*, Euro 3102. Copenhagen: WHO Regional Office for Europe, 1970.
 64. Bergman U. Pharmacoepidemiological perspectives. *J Clin Epidemiol* 1992; **45**: 313–7.
 65. Strom BL, Carson JL, Halpern AC, *et al*. Using a claims database to investigate drug-induced Stevens–Johnson syndrome. *Stat Med* 1991; **10**: 565–76.
 66. Anonymous. Panel III. Intervention: the use of drug utilization review outcomes. *Clin Pharmacol Ther* 1991; **50**: 633–5.
 67. Grimsson A. Drug utilization research. *Pharm Int* 1980; December: 245–7.
 68. Bergman U, Elmes P, Halse M, *et al*. The measurement of drug consumption: drugs for diabetes in Northern Ireland, Norway and Sweden. *Eur J Clin Pharmacol* 1975; **8**: 83–9.
 69. WHO Drug Utilization Research Group (DURG). Validation of observed differences in the utilization of antihypertensive and antidiabetic drugs in Northern Ireland, Norway and Sweden. *Eur J Clin Pharmacol* 1985; **29**: 1–8.
 70. WHO Drug Utilization Research Group (DURG). Therapeutic traditions in Northern Ireland, Norway and Sweden. I. Diabetes. *Eur J Clin Pharmacol* 1986; **30**: 513–9.
 71. WHO Drug Utilization Research Group (DURG). Therapeutic traditions in Northern Ireland, Norway and Sweden. II. Hypertension. *Eur J Clin Pharmacol* 1986; **30**: 521–5.
 72. Stålhammar J, Bergman U, Boman K, Dahlen M, *et al*. Metabolic control in diabetic subjects in three Swedish areas with high, medium, and low sales of antidiabetic drugs. *Diabetes Care* 1991; **14**: 12–9.
 73. Kennedy D. *Assessing National Drug Utilization in the United States*, WHO Drug Utilization Research Group Meeting. Washington, DC, 1983.
 74. Stolley PD, Lasagna L. Prescribing patterns of physicians. *J Chron Dis* 1969; **22**: 395–405.
 75. Baum C, Kennedy DL, Forbes MB, Jones JK. Drug use in the United States in 1981. *J Am Med Assoc* 1984; **251**: 1293–7.
 76. Baum C, Kennedy DL, Forbes MB, Jones JK. Drug use and expenditures in 1982. *J Am Med Assoc* 1985; **253**: 382–6.
 77. Hemminki E. Review of literature on the factors affecting drug prescribing. *Soc Sci Med* 1975; **9**: 111–15.
 78. Christensen DB, Bush PJ. Drug prescribing: Patterns, problems and proposals. *Soc Sci Med* 1981; **15**: 343–55.
 79. Soumerai SB, Avorn J. Efficacy and cost-containment in hospital pharmacotherapy: State of the art and future directions. *Milbank Mem Fund Q* 1984; **62**: 447–74.
 80. 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.
 81. Anderson GM, Lexchin J. Strategies for improving prescribing practice. *Can Med Assoc J* 1996; **154**: 1013–7.
 82. Swedish Drug Statistics (Svensk läkemedelsstatistik) 1997. Stockholm: Apoteket, 1998.
 83. Øydvín K, ed. *The drug consumption in Norway 1993–1997*. Oslo: Norwegian Medicinal Depot, 1998.
 84. Finnish Statistics on Medicines (Suomen laaketilasto) 1996. Helsinki, National Agency for Medicines and Social Insurance Institution, 1997.
 85. Gómez LA, Lobato CI. Evolución de las bases de datos de medicamentos del Ministerio de Sanidad y Consumo. *Inf Ter Seg Soc* 1987; **11**: 229.
 86. Avorn J, Soumerai SB. Use of computer-based Medicaid drug data to analyze and correct inappropriate medication use. *J Med Syst* 1982; **6**: 377–8.
 87. Wessling A. Continuous recording of drug prescribing in Sweden 1974–1983. Methods and examples of utilization of data. *Eur J Clin Pharmacol* 1987; **33**: 7–13.

88. Boëthius G, Wiman F. Recording of drug prescriptions in the county of Jämtland, Sweden. I. Methodological aspects. *Eur J Clin Pharmacol* 1977; **12**: 31–5.
89. Boëthius G. Approaches to assessing the rationality of drug usage in a developed country. *Acta Med Scand Suppl* 1988; **721**: 21–6.
90. Hallas J, Gaist D, Bjerrum L. The waiting time distribution as a graphical approach to epidemiologic measures of drug utilization. *Epidemiology* 1997; **8**: 666–70.
91. Steffensen FH, Kristensen K, Ejlersen E, et al. Major haemorrhagic complications during oral anticoagulant therapy in a Danish population-based cohort. *J Intern Med* 1997; **242**: 497–503.
92. Rosholm J-U, Gram LF, Isacsson G, Hallas J, Bergman U. Changes in the pattern of antidepressants use upon the introduction of the new antidepressants: a prescription database study. *Eur J Clin Pharmacol* 1997; **52**: 205–9.
93. Gaist D, Hallas J, Hansen N-CG, Gram LF. Are young adults with asthma treated sufficiently with inhaled steroids? A population-based study of prescription data from 1991 and 1994. *Br J Clin Pharmacol* 1996; **41**: 2885–9.
94. Gaist D, Tsiropoulos I, Sindrup SH, et al. Inappropriate use of sumatriptan: population based register and interview study. *Br Med J* 1998; **316**: 1352–3.
95. Steffensen FH, Kristensen K, Ejlersen E, et al. Major hemorrhagic complications during oral anticoagulant therapy in a Danish population-based cohort. *J Intern Med* 1997; **242**: 4997–503.
96. Olesen C, Steffensen FH, Sorensen HT, et al. Low use of long-term hormone replacement therapy in Denmark: a 5-year population-based survey. *Br J Clin Pharmacol* 1999; **47**: 323–8.
97. Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: How? *Am J Med* 1984; **77**: 703–8.
98. Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: Why? *Am J Med* 1985; **78**: 475–80.
99. Bergman U, Dahlström M. Anxiety and antianxiety drug use in Sweden. In: Strandberg K, Beermann B, Lonnerholm G, eds. Pharmacological treatment of anxiety. Uppsala, National Board of Health and Welfare Drug Information Committee, 1988, 63–74
100. Sanz E, Bergman U, Dahlström M. Pediatric drug prescribing—a comparison between Tenerife (Canary Islands, Spain) and Sweden. *Eur J Clin Pharmacol* 1989; **37**: 65–8.
101. Isacson D, Smedby B. Patterns of psychotropic drug use in a Swedish community. *Scand J Prim Health Care* 1988; **6**: 51–8.
102. Isacson D, Ståhlhammar J. Prescription drug use among diabetics—a population study. *J Chron Dis* 1987; **40**: 651–60.
103. Isacson D, Carsjö K, Bergman U, Blackburn JL. Long-term use, mortality and migration among benzodiazepine users in a Swedish community: an eight year follow-up. *J Clin Epidemiol* 1992; **45**: 429–36.
104. Blackburn JL. The use of automated databases in North America. In: Crommelin DJA, Midha KK, eds. *Topics in Pharmaceutical Sciences 1991*, proceedings of the 51st International Congress of Pharmaceutical Sciences of FIP, Washington, DC, 1991. Stuttgart: Medpharm, 1992, 405.
105. Strom BL, Carson JL, Morse ML, LeRoy AA. The computerized on-line Medicaid pharmaceutical analysis and surveillance system: a new resource for postmarketing drug surveillance. *Clin Pharmacol Ther* 1985; **38**: 359–64.
106. Strom BL, Morse ML. Use of computerized data bases to survey drug utilization in relation to diagnoses. *Acta Med Scand Suppl* 1988; **721**: 13–20.
107. World Health Organization. *How to investigate drug use in health facilities: selected drug use indicators*, WHO/DAP/93.1. Geneva: World Health Organization, 1993.
108. Hogerzeil VH, Bimo, Ross-Degnan D, et al. Field tests for rational drug use in twelve developing countries. *Lancet* 1993; **343**: 1408–10.
109. Sunartono. From research to action: the Gunungkidul experience. *Essential Drugs Monitor* 1995; **2**: 21–2.
110. Santoso B, Suryawati S, Prawitasari JE, Ross-Degnan D. Small group intervention vs formal seminar for improving appropriate drug use. *Soc Sci Med* 1996; **42**: 1163–8.
111. Ofori-Adjei D, Arhinful DK. Effect of training on the clinical management of malaria by medical assistants in Ghana. *Soc Sci Med* 1996; **42**: 1169–76.
112. Hadiyono JEP, Suryawati S, Danu SS, et al. Interactional group discussion: results of a controlled trial using a behavioral intervention to reduce the use of injections in public health facilities. *Soc Sci Med* 1996; **42**: 1177–83.
113. Baksaas-Aasen I, Lunde PKM, Halse M, et al. Drug dose statistics. List of defined doses for drugs registered in Norway. Oslo, Norwegian Medicinal Depot, 1975.
114. Nordic Council on Medicines. *Nordic Statistics on Medicines 1993–95*, NLN Publication 43. Uppsala, Nordic Council on Medicines, 1996.
115. Anonymous. Guidelines for ATC classification and DDD assignment, 1st edn. Oslo: WHO Collaborating Centre for Drug Statistics, 1996.
116. Bergman U, Christenson I, Janson B, Wiholm BE. Auditing hospital drug utilisation by means of

- defined daily doses per bed-days. A methodological study. *Eur J Clin Pharmacol* 1980; **17**: 183–7.
117. Hekster YA, Vree TB, Goris RJ, Boerema JB. The defined daily dose per 100 bed-days as a unit of comparison and a parameter for studying antimicrobial drug in a university hospital. A retrospective study of the effects of guidelines and audit on antimicrobial drug use. *J Clin Hosp Pharm* 1982; **7**: 251–60.
 118. Kennedy DL, Forbes MB, Baum C, *et al.* Antibiotic use in US hospitals in 1981. *Am J Hosp Pharm* 1983; **40**: 797.
 119. Bergman U, Boman G, Wiholm BE. Epidemiology of adverse drug reactions to phenformin and metformin. *Br Med J* 1978; **2**: 464–6.
 120. Sjöqvist F. Drug utilization. In Tuomisto J, Paasonen MK, eds. *Proceedings of the Sixth International Congress of Pharmacology: Vol. 5. Clinical pharmacology*. Helsinki: Finland, Finnish Pharmacological Society, 1975, 39.
 121. Lee D, Chaves Matamoros A, Mora Duarte J. Changing patterns of antibiotic utilization patterns in Costa Rica. *APUA Newslett* 1991; **9**: 7.
 122. Laporte JR, Capellá D, Gisbert R, *et al.* The utilization of sedative–hypnotic drugs in Spain. In: Tognoni G, Bellantuono C, Lader M, eds. *Epidemiological impact of psychotropic drugs*. Amsterdam: Elsevier–North-Holland, 1981, 37.
 123. Tawara K, Nakai K, Sudou I, *et al.* Record-linked drug utilization data in a Japanese university hospital: Sedatives, hypnotics and antianxiety drugs. *Acta Pharmacol Toxicol Suppl* 1986; **5**: abstract 1246.
 124. Truter I, Wiseman K, Kotze TT. The defined daily dose as a measure of drug consumption in South Africa. A preliminary study. *S Afr Med J* 1996; **86**: 675–9.
 125. Busto U, Isaac P, Lanctot K, *et al.* Psychotropic drug utilization in Canada: 1978–1984. *Acta Pharmacol Toxicol Suppl* 1986; **5**: abstract 1255.
 126. Lee D, García LR, Mewa D, *et al.* Piroxicam utilization in Panama. *Acta Pharmacol Toxicol Suppl* 1986; **5**: abstract 1263.
 127. Col NF, O'Connor RW. Estimating worldwide current antibiotic usage: Report of Task Force 1. *Rev Infect Dis* 1987; **9**: S232–43.
 128. Kiiwet RA, Bergman U, Rootslane L, *et al.* Drug use in Estonia in 1994–1995: a follow-up from 1998 and comparison with two Nordic countries. *Eur J Clin Pharmacol* 1998; **54**: 119–24.
 129. Commonwealth Department of Health and Family Services. *Australian Statistics on Medicines 1997*. Canberra: Commonwealth of Australia, 1998.
 130. Stanulovic M, Milosev M, Jakovlojevic V, Roncevic N. Epidemiological evaluation of anti-infective drug prescribing for children in outpatient practice. *Dev Pharmacol Ther* 1987; **10**: 278–91.
 131. Bergman U, Sjöqvist F. Measurement of drug utilization in Sweden: methodological and clinical implications. *Acta Med Scand* 1984; **105**: 15–22.
 132. Bergman U. Utilization of antidiabetic drugs in the island of Gotland, Sweden: Agreement between wholesale figures and prescription data. *Eur J Clin Pharmacol* 1978; **14**: 213–20.
 133. Bergman U, Dahlström M. Användningen av blodtryckssänkande läkemedel i Sverige Use of antihypertensive agents in Sweden. In: Berglund G, ed. *Hypertoni 87*. Mölndal: Lindgren and Söner, 1988, 22–32.
 134. Bergman U, Dahlström M. Benzodiazepine utilization in Sweden and other Nordic countries. In: Lader MH, Davies HC, eds. *Drug treatment of neurotic disorders*. Edinburgh: Churchill Livingstone, 1986, 43–52.
 135. Juergens JP, Bergman U, Baum C, *et al.* Use of benzodiazepines in the USA and Sweden. *Drug Information Association Annual Meeting*, Washington, DC, 1986.
 136. Isacson G, Holmberg P, Wasserman D, Bergman U. Use of antidepressants among people committing suicide in Sweden. *Br Med J* 1994; **308**: 506–9.
 137. Isacson G, Holmberg P, Druid H, Bergman U. The utilization of antidepressants—a key issue in the prevention of suicide. An analysis of 5281 suicides in Sweden 1992–1994. *Acta Psychiatr Scand* 1997; **96**: 94–100.
 138. Anonymous. *Ira Lista de Medicamentos Esenciales*. Venezuela: Ministerio de Sanidad y Asistencia Social, 1990.
 139. Anonymous. *Lista Nacional de Medicamentos*. Guatemala: Ministerio de Salud Pública y Asistencia Social, 1993.
 140. Comisión de Medicamentos. *Lista Oficial de Medicamentos*. Panama: Caja de Seguro Social, 1997.
 141. The National Essential Drugs Committee. *Standard Treatment Guidelines and Essential Drugs List for Primary Health Care*, 2nd edn. Pretoria: National Department of Health, 1998.
 142. Pahor M, Chrischilles EA, Guralnik JM, *et al.* Drug data coding and analysis in epidemiologic studies. *Eur J Epidemiol* 1994; **10**: 405–11.
 143. Anonymous. Appendix IV: VA medication classification system. In: *USP Dispensing Information: Vol. I. Drug information for the health care professional*. Englewood, CO: Micromedex, 1999, 3045–62.
 144. Avorn J, Soumerai SB. Improving drug-therapy decisions through educational outreach: A randomized controlled trial of academically based “detailing.” *New Engl J Med* 1983; **308**: 1457–63.
 145. Schaffner W, Ray WA, Federspiel CF, Miller WO. Improving antibiotic prescribing in office practice.

- A controlled trial of three educational methods. *J Am Med Assoc* 1983; **250**: 1728–32.
146. Ray WA, Schaffner W, Federspiel CF. Persistence of improvement in antibiotic prescribing in office practice. *J Am Med Assoc* 1985; **253**: 1774–6.
147. Morse ML, LeRoy AA, Gaylord TA, *et al.* Reducing drug therapy-induced hospitalization: impact of drug utilization review. *Drug Inf J* 1982; **16**: 199–202.
148. Groves R. Therapeutic drug-use review for the Florida Medicaid program. *Am J Hosp Pharm* 1985; **42**: 316–9.
149. Soumerai SB, Avorn J, Ross-Degnan D, Gortmaker S. Payment restrictions for prescription drugs under Medicaid. *New Engl J Med* 1987; **317**: 550–6.
150. O'Connell DL, Henry D, Tomlins R. Randomised controlled trial of feedback on general practitioners prescribing in Australia. *Br Med J* 1999; **318**: 507–11.
151. Anderson GM, Spitzer WO, Weinstein MC, *et al.* Benefits, risks, and costs of prescription drugs: a scientific basis for evaluating policy options. *Clin Pharmacol Ther* 1990; **48**: 111–9.
152. Schiff GD, Rucker TD. Computerized prescribing: building the electronic infrastructure for better usage. *J Am Med Assoc* 1998; **279**: 1024–9.
153. Von Ferber C. Confidentiality Working Group questionnaire. *EURO DURG Bull* 1998; **3**: 3–4.
154. Lewis MA. ISPE: present, past and future. *PharmacoEpidemiol Newslett* 1997; **11**: 4.
155. Guess HA, Goldsmith C, Henry D, Strom BL. Pharmacoepidemiology as a focus for clinical epidemiology in developing countries. *J Clin Epidemiol* 1991; **44**: 101–5S.
156. Haaïjer-Ruskamp FM, Bergman U. Rational drug use in Europe—challenges for the 21st century. Report from the 1st meeting of EURO DURG, the European Drug Utilization Research Group. *Eur J Clin Pharmacol* 1997; **52**: 1–8.
157. Ross-Degnan D, Laing R, Quick J, *et al.* A strategy for promoting improved pharmaceutical use: the International Network for Rational Use of Drugs. *Soc Sci Med* 1992; **35**: 1329–41.
158. Lee D. The International Network for Rational Use of Drugs. *Essential Drugs Monitor* 1997; **24**: 7–8.

30

Evaluating and Improving Physician Prescribing

STEPHEN B. SOUMERAI

Department of Ambulatory Care and Prevention, Harvard Medical School and
Harvard Pilgrim Health Care, Boston, MA, USA

SUMIT MAJUMDAR

Division of General Internal Medicine, University of Alberta, Edmonton, Alberta, Canada

HELENE LEVENS LIPTON

Department of Clinical Pharmacy, School of Pharmacy and Institute for Health Policy Studies,
School of Medicine, University of California at San Francisco, San Francisco, CA, USA

Research and clinical practice may be on parallel tracks headed in the same direction, but in contact only through rotting ties.¹

INTRODUCTION

The broad purposes of pharmacoepidemiology are to advance our knowledge of the risks and benefits of medication use in real-world populations, and to foster improved prescribing and patient health outcomes. If, however, physicians and other health practitioners fail to update their knowledge and practice in response to new and clinically important data on the outcomes of specific prescribing patterns, then the “fruits” of pharmacoepidemiology research may have little impact on clinical practice.

It is for these reasons that a new discipline in the fields of health services research and clinical decision

making has grown rapidly in importance—the science of assessing and improving clinical practices. The rapid growth of this new field, fueled by increasing research support from the National Institutes of Health and the Agency for Healthcare Research and Quality, is based on the recognition that passive knowledge dissemination (e.g., distributing practice guidelines) is generally insufficient to improve clinical practices without supplemental behavioral change interventions based on relevant theories of diffusion of innovations, persuasive communications, and adult learning theory.^{2–6}

This chapter reviews some of these developments as they relate to medication use; defines several types of drug prescribing problems; discusses several thorny methodologic problems in this literature; reviews existing pharmacoepidemiologic and other evidence on the effectiveness of common prescribing-improvement interventions;

and concludes with a discussion of future research needs. For a more detailed and comprehensive examination of the literature on prescribing education, the role of the clinical pharmacist as a change agent, and managerial strategies for use in various settings, the reader is advised to consult several previous works published elsewhere.⁷⁻¹⁸ Portions of this chapter are derived from this body of work; in addition, we conducted computerized literature searches (published through 1998), consulted personal files, hand-searched references, and consulted The Cochrane Collaboration on Effective Professional Practice, a rigorous and continuously updated registry and synthesis of available evidence on studies of interventions to change physician behaviors.¹⁹ A substantial, if uneven, literature also exists on the intended and unintended impacts of cost sharing and other reimbursement restrictions designed to control drug expenditures. Review articles,^{20,21} a synthesis of two decades of research,²² and two additional studies in this area have recently been published,^{23,24} and, because it is beyond the scope of this chapter, will not be reviewed here.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

There is little doubt that the importance of suboptimal prescribing practice (both under- and overuse) vastly outweighs the costs of medications themselves^{25,26} (see Chapter 35). Drug therapies are the most common treatments in medical practice; their potential for both alleviating and causing illness are illustrated throughout this book. The types of prescribing error are at least as varied as their causes, and include

- use of toxic or addictive drugs when safer agents are available (e.g., barbiturates instead of benzodiazepines),
- use of drug therapy when no therapy is required (e.g., antibiotics for viral respiratory infections),
- use of an ineffective drug for a given indication (e.g., cerebral vasodilators for senile dementia),
- use of a costly drug when a less expensive preparation would be just as effective (e.g., newer calcium channel blockers, instead of effective and inexpensive thiazide diuretics, for uncomplicated hypertension),
- under- or excessive use of effective agents (e.g., too high dosages of benzodiazepines for the elderly),
- failure to discontinue therapy when the drug is no longer needed (e.g., use of H_2 -blockers for four or more months),
- failure to introduce new and effective drugs into practice (e.g., inhaled corticosteroids for asthma),
- failure to prescribe necessary drug therapies (e.g., use of β -blockers following acute myocardial infarction), and
- failure to achieve recommended therapeutic goals (e.g., LDL cholesterol levels below 100 mg/dL for the secondary prevention of myocardial infarction).

Specific illustrations of the above problem categories are ubiquitous in the drug utilization literature (although not always carefully documented). For example, propoxyphene, a toxic and abusable narcotic analgesic, is often prescribed for mild-to-moderate pain when other safer, more effective analgesics are available.^{27,28} In the hospital setting, numerous studies have documented that as much as 50% of antibiotic use is inappropriate, based on well established guidelines (e.g., wrong choice of agent, incorrect administration, drug not required). Furthermore, overuse of antibiotics may be increasing microbial resistance to these agents.²⁹ In the nursing home setting, abundant evidence exists that potent anti-psychotic drugs are often administered inappropriately to demented elderly patients, posing serious iatrogenic risks.^{30,31}

Because of the absence of diagnostic data in most published drug utilization research, and because of the emphasis on cost containment within drug utilization review programs, the existing literature may *underemphasize* the important problem of underuse of highly effective medications. For example, a study in a large US health maintenance organization found that nearly two-thirds of newly diagnosed hypertensive

patients (those with diastolic blood pressures of greater than 100 mm Hg) were not followed up for treatment 6–12 months after diagnosis.³² In addition, a 1990 study of 623 outpatients treated for acute myocardial infarction at the Yale-New Haven Hospital found that one-third of the patients meeting strict randomized clinical trial eligibility criteria for use of β -blockers did not even receive a trial of therapy—contrary to existing guidelines from the American College of Cardiology. As a result, they experienced a 20–40% higher post-myocardial infarction mortality rate.³³

Another recent study found that almost 60% of medical and surgical inpatients reported “horrible” or “excruciating” pain, despite the availability of safe and powerful analgesics which can eliminate virtually all pain in this setting.³⁴ The authors concluded that irrational fears of prescribing narcotic analgesics for acute pain were significant barriers to appropriate treatment.

Finally, recent studies have documented that strict reimbursement restrictions on medications among chronically ill elderly in Medicaid were associated with reductions in use of essential agents (e.g., insulin, cardiovascular drugs, and lithium) and subsequent increases in the risk of nursing home admission.^{35,36}

Why do these problems occur? Can a comprehensive theory of behavioral change provide the basis for programs designed to improve prescribing? Such an ideal model must be complex given the diversity of economic, organizational, educational, psychological, social, and informational influences on daily prescribing practices. Some of the factors responsible for suboptimal prescribing include the failure of clinicians to keep abreast of important new findings on the risks and benefits of medications; overpromotion of some drugs through pharmaceutical company advertising or sales representatives (i.e., “detailers”); underpromotion of highly effective, but non-profitable medications (e.g., aspirin); simple errors of omission;³⁷ negative attitudes towards issues of cost-effectiveness of medications; direct-to-consumer marketing strategies; patient and family demand for a particular agent, even when it is not scientifically substantiated;³⁸ physician

overreliance on clinical experience in opposition to scientific data;³⁸ a skepticism toward, and distrust of, the literature and academia among some community-based physicians;³⁹ the need to take some “definitive” therapeutic action, even when “watchful waiting” may be the most justifiable action under conditions of uncertainty; outdated therapeutic protocols; and influence from clinical opinion leaders or other health practitioners (e.g., clinical pharmacists). These diverse influences suggest the need for tailoring intervention strategies to the key factors influencing a given clinical behavior based on models of behavioral change. One promising model will be discussed in the section entitled “Currently Available Solutions.”

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY RESEARCH

Research on the impact of educational and administrative interventions to improve drug prescribing presents numerous methodological challenges. This section will review several of the most important methodological problems and suggested solutions, including internal validity, regression toward the mean, unit of analysis, logistical issues, ethical and legal problems, and the detection of effects on patient outcomes.

INTERNAL VALIDITY

As early as 1975, Gilbert *et al.* established that poorly controlled studies produce misleading estimates of the effects of a variety of social programs.⁴⁰ Many nonintervention factors can affect medication utilization over time, such as marketing campaigns, mass media, state or federal regulatory policies, seasonal effects, changing in staffing of health care organizations, changes in eligibility for insurance programs, shifting demographics, etc. Because randomized clinical 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 comparison group) should be used instead of weak one-group post-only or pre-post designs that do not generally permit causal inferences.

Interrupted time-series designs include multiple observations (often ten or more) of study populations before and after intervention. Such designs permit investigators to control for pre-intervention secular changes in study outcomes and to estimate the size and statistical significance of sudden changes in the level or slope of the time series occurring at initiation of the treatment. The availability of a comparison series collected from a similar, but unexposed, comparison group can further increase causal inferences if no simultaneous change in trend is observed for this group.⁴¹

Another popular design that can often lead to interpretable results is the *pre-post with comparison group design*. This design includes a single observation both before and after treatment in a 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. Although this design controls for many threats to the validity of causal inferences (e.g., due to the effects of testing or maturation), it cannot control for unknown factors (e.g., a regulatory policy) which might result in pre-intervention differences in trends between study and comparison groups.

The weakest, and not uncommon, design is the *one-group, post-only design*, which consists of making only one observation on a single group which has already been exposed to a treatment.

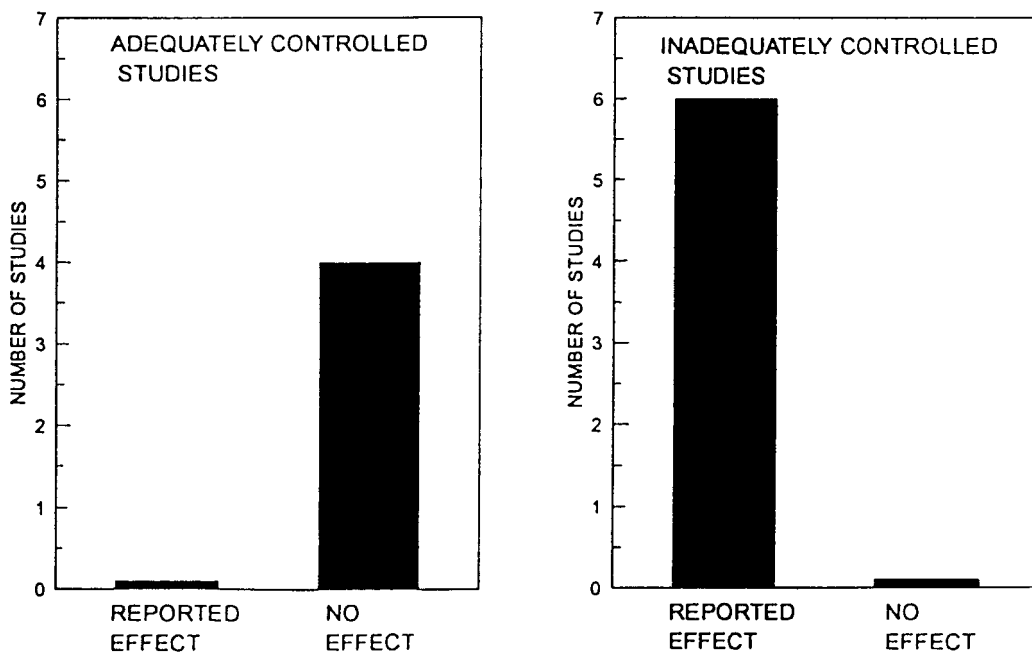


Figure 30.1. Reported effectiveness of dissemination of printed educational materials alone in well designed versus inadequately controlled studies. Reprinted with permission from the *Milbank Quarterly*.⁷

The *one-group pre–post design* merely adds a single pre-intervention observation to the previous design. As described below, such weak designs are unlikely to produce valid or reliable estimates of the effects of educational or managerial interventions. Unfortunately, however, 60% of 76 studies designed to improve drug prescribing in primary care and inpatient settings used the weakest nonexperimental designs.⁷

Inadequately controlled studies may exaggerate the effectiveness of many interventions to improve prescribing. For example, as shown in Figure 30.1, inadequately controlled studies of the dissemination of print-only materials used alone (right side) have all reported positive effects on behavior, while well controlled studies of such strategies (left side) all reported small or non-existent changes in behavior. The “success” of uncontrolled studies is often due to the attribution of pre-existing trends in practice patterns to the studied intervention.

The potential bias of failing to account for prior trends was recently demonstrated by examining naturally occurring trends in utilization of 23 specific categories of substitute medications whose utilization averaged at least one prescription per 1000 enrollees per month in a 4 year study of 390 000 enrollees in the New Jersey Medicaid program.⁴² The results indicated that 50% of the estimated 1-year percent changes in prescriptions per 1000 enrollees exceeded +20.3% or –10.8% of baseline levels. It is interesting to note that the size of many of the “effects” reported in the drug utilization/intervention literature⁷ are similar to these natural fluctuations, suggesting that changes in drug use attributed to such interventions could merely reflect these underlying trends.

The above findings provide further support for more widespread application of RCTs or, when RCTs are not feasible, time-series and comparison series designs to evaluate whether suddenly introduced interventions are associated with corresponding changes in the level or slope of the utilization series, after controlling for prior trends (see references 23, 27, 36 for examples). If the collection of time-series data is not feasible,

investigators may consider utilizing pre–post with comparison group designs, which also control for most threats of history, as described in respected texts on intervention research design.⁴¹

REGRESSION TOWARD THE MEAN

Regression toward the mean, the tendency for observations on populations, selected on the basis of exceeding a predetermined threshold level, to approach the mean on subsequent observations, is a common and insidious problem in much of the drug utilization literature. For example, the most common Medicaid drug utilization review programs typically screen utilization data and eligibility files for possible co-occurrences of two interacting medications, or higher than recommended dosages for individual drugs. After case-by-case review by expert committees, letters are written to responsible physicians questioning the practice and asking for written responses. Unfortunately, however, the only published research evaluating this methodology utilized poorly controlled designs that are unable to control for regression to the mean. For example, in one often cited drug utilization review study,⁴³ 50% of prescribing problems were absent several months after letters were sent, suggesting to the noncritical reader that the program was effective. However, it is equally plausible that the offending medications were withdrawn because the patients’ conditions improved or because the physicians detected the error on their own.

The likelihood that all screening algorithms employed in drug utilization review programs are subject to regression toward the mean argues strongly for the need to conduct randomized controlled experiments and well controlled quasi-experiments (e.g., time series with comparison series) to justify the efficiency and effectiveness of these programs *before* they become a routine part of private and public quality-improvement methods. If regression effects are unavoidable—for example, due to selection of at-risk (high-use) populations—investigators may consider including a “wash-out”

period after selection and before pre- and post-intervention observations.³⁵

UNIT OF ANALYSIS

A common methodological problem in studies of physician behavior is the incorrect use of the patient as the unit of analysis. Such a practice violates basic statistical assumptions of independence because prescribing behaviors for individual patients are likely to be correlated within each physician's practice. This often leads to exaggerated significance levels when the correct unit of analysis is the physician or health care facility. As a result, interventions may appear to lead to "statistically significant" improvements in prescribing practices when in fact no such claim is warranted.

A recent review of articles on physicians' patient care behavior found that 70% of 54 articles incorrectly analyzed the data using the patient as the unit of analysis;⁴⁴ among 19 reviewed studies of medication prescribing, 58% used the incorrect unit of analysis.

The simplest, although sometimes overly conservative, solution to the problem of incorrect unit of analysis is to analyze data by facility or physician. Alternatively, new methods for analyzing clustered data are also becoming increasingly available; such models can control for clustering of observations at the patient, physician, and facility levels.⁴⁵ Such models allow aggregation at the patient level by controlling for correlation between patients cared for by the same provider or facility. The resulting significance levels for differences in prescribing rates between study and control groups are more conservative than assuming no intraclass correlation, but are greater than the most conservative methods of analyzing at the provider or facility level.

LOGISTICAL ISSUES

While continuity of care is a goal in most settings, many patients, particularly those treated within academic medical centers, see multiple primary providers over time. For example, patients treated by residents may be reassigned to other residents at

the end of the academic year. Providers may go on extended leave and transfer cases to other clinicians. Patients themselves may choose another primary care provider. In addition, many patients develop ongoing relationships with specialists as particular problems develop and are resolved.

While these changes may or may not improve patients' care, they almost always complicate and sometimes weaken research conducted in a clinical setting. Particularly in settings where providers may be assigned to both "intervention" and "control" patients, contamination problems are difficult to avoid. Even when interventions can be focused effectively on the intended patients or providers, informal communication among providers can lead to contaminated effects, thereby decreasing the likelihood of detecting significant changes.

Fortunately, solutions to the above problems exist. First, investigators should identify through baseline interviews and organizational records the extent to which patients are cared for by multiple providers, and the patterns of consultations and referrals between caregivers within and between facilities. If randomization of clinicians is likely to lead to contamination of controls, or if patient-provider pairs are frequently broken, the entire facility or subunit (e.g., primary care center) should be assigned to the same study group. For instance, a recent quality-improvement intervention randomized 37 hospitals in one state to intervention or control status.⁴⁶ However, when this strategy is not feasible, because it results in a small sample of facilities and inadequate statistical power, investigators are encouraged to collect data on medication use during multiple observation periods both before and after the intervention, and to use time-series regression methods that can often detect modest changes in utilization levels after as few as 7–12 months.^{10, 35, 47}

ETHICAL AND LEGAL PROBLEMS HINDERING THE IMPLEMENTATION OF RANDOMIZED CLINICAL TRIALS

Adequate control groups are essential for rigorous evaluation of results. Yet it has been argued that there are ethical and legal problems related to

withholding interventions designed to improve drug-prescribing practices. This is especially true in government-funded programs such as Medicaid.

This argument assumes that the proposed interventions are known to be beneficial. In fact, the efficacy of many programs to improve drug utilization is the very question that should be under investigation. In those rare instances in which the intervention has shown unusual promise in similar populations, the application of randomized clinical trials may be inappropriate. In such cases, alternative research designs should be considered. Feasible design alternatives are quasi-experimental designs such as interrupted time-series analysis, or staged implementation in which the control population (or regions) receive the intervention after comparative data have been collected.^{22,41}

DETECTING EFFECTS ON PATIENT OUTCOMES

While a number of studies have demonstrated positive effects of various interventions on prescribing practices, almost no large well controlled studies have linked such changes in prescribing to improved patient outcomes. Recently, under the auspices of the Health Care Financing Administration, Marciniak *et al.* conducted a controlled trial of guideline dissemination and feedback by peer-review organizations on seven quality (i.e., process) indicators for acute myocardial infarction care in Medicare patients in four states.⁴⁸ This was a before (1992–93) and after (1995–96) intervention study, with a post-only comparison to a random national sample of patients from the other states. Almost 24 000 patient records were abstracted. Performance on all quality indicators improved in the intervention states compared to baseline; however, only the use of aspirin and β -blockers and counseling for smoking cessation were significantly greater than in the control states. An important strength of this study, beyond its size and scope, was an analysis of patient outcomes, namely mortality. There was no difference in mortality between intervention and control states in 1992–93, but after the intervention and consistent with documented improvements in process, mortality was approximately 1% lower (1 year mortality

30.4% versus 31.4% in the control states, $p = 0.004$) in the intervention states. Bearing in mind certain important threats to validity (e.g., no baseline measurement of process indicators in control states, and possible lack of comparability between intervention and control states), this is one of the few studies that suggest a link between improvements in process and patient outcomes.

These results, and those of other studies,³¹ underline the difficulty of demonstrating statistically significant changes in patient outcomes in response to intervention. Explanations for the dissociation between improvements in prescribing and better patient outcomes include (i) available clinical outcome measures may not be sensitive to the kinds of patient outcome that might be affected by introduction or withdrawal of medications, (ii) changes in physician prescribing may lead to little or no change in patients' health status if patients do not adhere to the recommended regimens, and (iii) many medical therapies require months to years of continued compliance before clinical benefits become apparent.

Because of the above problems, sample sizes may need to be enormous to detect modest changes in patient outcomes (see Chapter 3 for a discussion of methods for determining statistical power). These problems are much less severe in standard drug trials because of experimenter control over the major independent variable—exposure to medications. However, process outcomes (e.g., use of recommended medications for acute myocardial infarction from evidence-based practice guidelines) may often be sensitive and appropriate measures of quality of care,⁴⁹ and improvements in process should not be dismissed outright as surrogate outcomes. They may be important in and of themselves, as long as the processes are a measure of proven effective therapy.

CURRENTLY AVAILABLE SOLUTIONS

CONCEPTUAL FRAMEWORK

A useful starting point for designing prescribing-improvement interventions 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 colleagues,⁵⁰ and proposes factors influencing three sequential stages of behavior change: predisposing, enabling, and reinforcing factors. Predisposing variables include such factors as awareness of a consensus guideline on appropriate use of a thrombolytic agent, knowledge of clinical relationships supporting such a guideline (e.g., major actions of thrombolytics in the artery), beliefs in the efficacy of treatment (e.g., probability of survival), attitudes or values associated with recommended behaviors, etc. However, while a mailed drug bulletin may predispose some physicians to new information (if they read it), behavior change may be impossible without new enabling skills (e.g., skills in administering a new therapy, or overcoming patient or family demand for unsubstantiated treatments). Once a new pattern of behavior is tried, multiple and positive reinforcements (e.g., through peers, reminders, or positive feedback) may be necessary to establish fully the new behavior. A number of recent thoughtful reviews of the literature have come to a similar conclusion: multi-faceted interventions that encompass all stages of behavior change are most likely to improve physician prescribing.^{6, 7, 13, 15, 19, 51}

EMPIRICAL EVIDENCE ON THE EFFECTIVENESS OF PRESCRIBING-IMPROVEMENT INTERVENTIONS

Does existing empirical evidence on the effectiveness of alternative prescribing interventions provide any lessons on the key characteristics of successful approaches to this problem? Illustrative findings from several research syntheses will be used to evaluate the effectiveness of the most commonly studied or applied approaches. Because of severe biases introduced by uncontrolled designs which do not measure pre-existing trends in target drug use behaviors (see prior Methodologic Problems section), only studies utilizing adequate experimental or quasi-experimental research designs (e.g., pre-post with comparison group and time-series designs) are discussed.

DISSEMINATION OF EDUCATIONAL MATERIALS AND CLINICAL PRACTICE GUIDELINES

Distributing printed educational materials aimed at improving prescribing practice remains the most ubiquitous form of prescribing education in the industrialized world. While the most sophisticated materials may incorporate visually arresting graphs, illustrations, and headlines to convey important behavioral and educational messages, such a strategy rests on assumptions that physicians will be exposed to the information, and that such rational information will be sufficiently persuasive to change clinical practices. Unfortunately, several reviews provide consistent evidence that use of disseminated educational materials *alone* (such as drug bulletins, self-education curricula, objective, graphically illustrated “un-advertisements,” or other professionally prepared educational brochures) may affect some of the predisposing variables in the change process (e.g., knowledge or attitudes), but has little or no effect on actual prescribing practice.

A study of the effect of warning letters mailed to 200 000 physicians who were high prescribers of zomepirac sodium corroborate this previous literature.²⁸ As shown in Figure 30.2, the warning letters, which alerted these physicians to serious or fatal anaphylaxis associated with use of zomepirac, were not associated with any reduction in its use, especially in the face of stronger face-to-face marketing campaigns which may have counteracted the warning messages. Several previous controlled trials in five states also failed to document the effects of well illustrated brochures on use of oral cephalosporins, contraindicated antibiotics, propoxyphene (a marginally effective analgesic), and ineffective agents for senility and claudication.⁵²⁻⁵⁴ A distinct subset of educational materials are clinical practice guidelines. Although primarily educational in nature, they are also a codification of current best practice, and are intended to improve quality and decrease costs by decreasing variations in practice. However, faith in the simple act of guideline dissemination presupposes that information alone, regardless of how reliable or how well referenced, can change

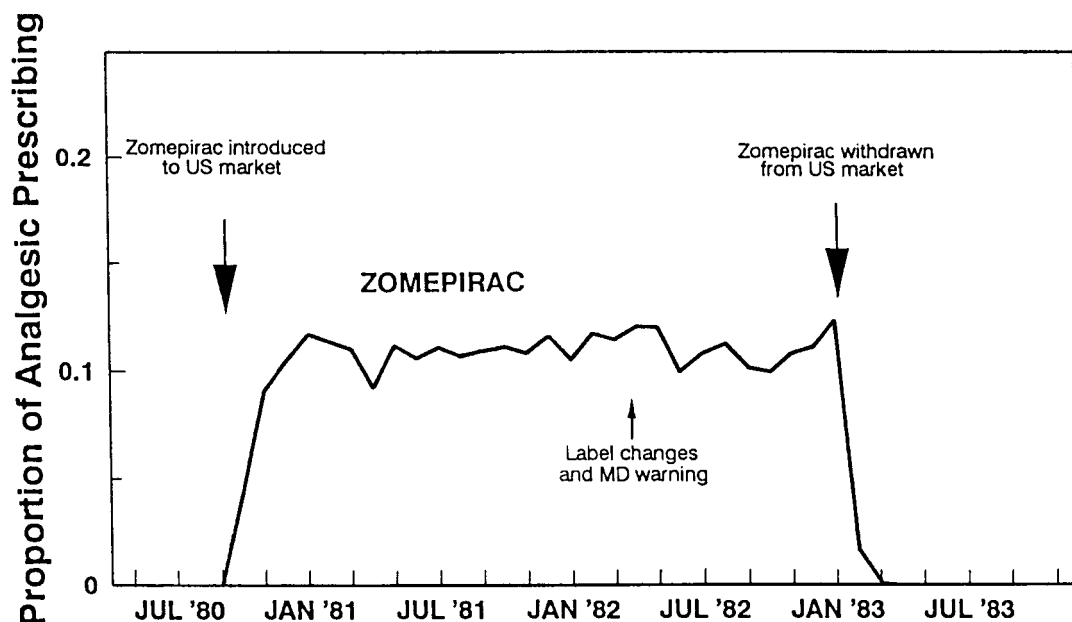


Figure 30.2. Time series of zomepirac prescribing in Medicaid (as a proportion of total analgesic prescriptions) among 260 primary care physicians before and after warnings concerning possible severe adverse effects. Reprinted with permission from *JAMA*.²⁸

behavior. When rigorously studied, guideline dissemination *alone* has not significantly influenced prescribing behavior or other clinical practices.^{6, 15, 17, 55, 56}

In general, simple dissemination of educational materials does not appear to be effective by itself in altering prescribing patterns, but these materials may provide a necessary “predisposing” foundation for other “enabling” and “reinforcing” strategies.

MULTIMEDIA WARNING CAMPAIGNS

Occasionally, the discovery of important adverse effects of marketed drugs is accompanied by mailed educational materials to physicians as part of a broader warning campaign involving the medical and popular press, newspapers, television, and radio. When the adverse effects are severe and preventable, alternative agents exist, and the messages are simple enough to convey in mass communications, such multimedia campaigns may sometimes be effective in changing prescribing

patterns in large populations. Previous examples include marked reductions in the use of chloramphenicol (aplastic anemia)⁵⁷ and much smaller reductions in the use of calcium channel blockers (myocardial infarction) in response to widespread media warnings.⁵⁸

Figure 30.3 provides data from a US study suggesting that widespread reporting of the risk of Reye’s syndrome associated with pediatric aspirin use by the medical and lay press was associated with declines in Reye’s syndrome. This media campaign was conducted after Reye’s syndrome was linked to aspirin use and several antecedent viral illnesses in several epidemiological studies.⁵⁹ The authors concluded, based on this and other studies, that mass media warnings may be effective in changing both consumer and physician behavior when the illness is severe or life-threatening, the behavioral message is simple, no or few barriers to alternative behaviors (e.g., acetaminophen versus aspirin) are present, and the campaign is comprehensive, involving both health professionals and consumers.

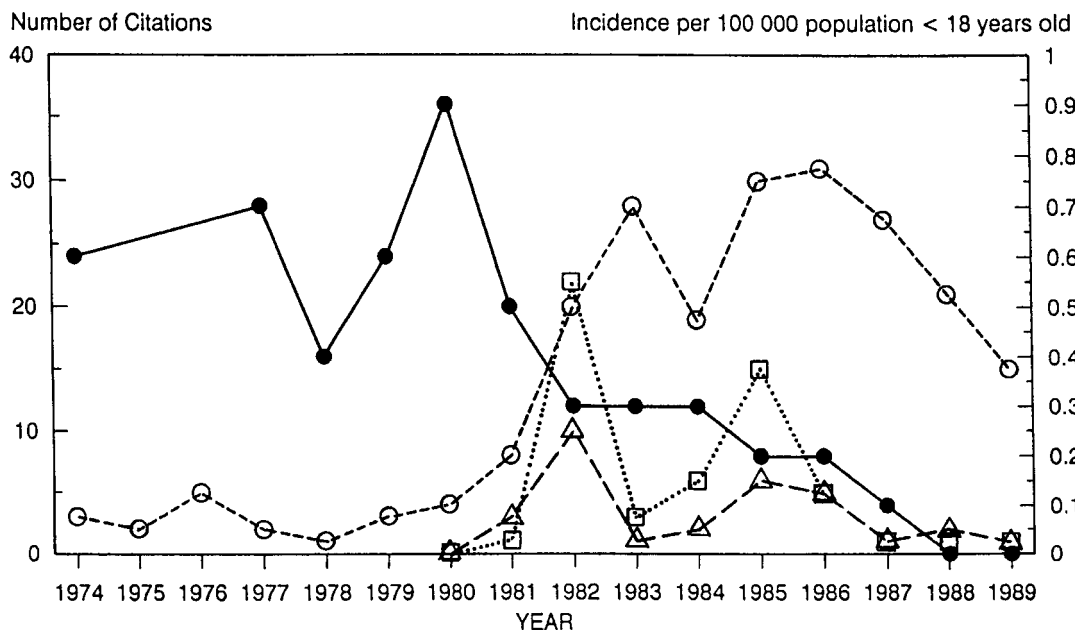


Figure 30.3. Trend in number of medical and lay press citations on aspirin and Reye's Syndrome, and in incidence of Reye's syndrome among children. Newspaper index limited to four continuously reporting national newspapers described in text. Reprinted with permission from the *Milbank Quarterly*.⁵⁹ Incidence of Reye's syndrome, ●; medical literature citations, ○; newspaper citations, □; magazine citations, △.

GROUP EDUCATION

Although rounds, seminars, and other group didactic educational programs are among the most universal methods for prescribing education, controlled studies of this approach are almost nonexistent in the literature, especially in nonteaching settings. Nevertheless, small group discussions conducted by clinical leaders in academic primary care settings have been shown to improve use of antibiotics and agents for hypertension treatment and control.^{60,61} These successful approaches have included reviews of patient records to establish the need for change and participatory methods based on adult learning theory. Traditional large group, didactic continuing medical education seminars have not been as successful, by themselves, in improving physician performance for other medical technologies.

PROFILING, AUDIT, AND FEEDBACK

During the last 20 years, a new approach to improving physician performance has become

increasingly popular—feedback of prescribing patterns to individuals or groups of physicians. A recent survey found more than half of all US physicians received clinical or economic feedback regarding their prescribing practices.¹⁶ While managers and health policy makers often assume that “feedback” is a unidimensional technique, its many variations have not been well defined or well studied.

One well studied form of feedback is patient-level medication profiles. It has frequently been hypothesized that simply making clinicians aware of all of the medications a patient may be prescribed might be an effective method for reducing use of excessive, duplicative, or interacting medications. The best controlled trials of this approach confirm that simply distributing such profiles, without explicit suggestions for changes in practices, has no detectable effect on prescribing practice.⁶²⁻⁶⁴ Likely reasons for the failure of this intuitively appealing approach include the following: (i) much of the generated information was probably clinically irrelevant; (ii) unsynthesized

and voluminous data may cause information “overload” and desensitization of busy clinicians; (iii) there was no provision of alternative measures to improve care; and (iv) the feedback was not derived from credible sources of information. This approach represents one of the few instances in which the volume of negative findings from methodologically rigorous studies strongly supports the exclusion of this approach from prescribing-improvement options and future research.

Other forms of feedback may compare practice patterns with peers or predetermined standards such as practice guidelines. The former is typified by interventions of peer-comparison feedback, while drug utilization review (DUR) programs typify the latter (DUR is discussed extensively in Chapter 31 and will not be covered here). A recent meta-analysis of 12 RCTs concluded that peer comparison feedback had a statistically significant, but clinically minimal, effect on prescribing or other physician behaviors.¹⁶ Further, the authors doubted that such programs were likely to offset the costs of the intervention, much less lead to cost savings.

In addition to the type or content of the feedback, a number of variables must be considered. Communication channels could be by letter, computer, or face-to-face encounter with a supervisor or colleague. The credibility of the source of the feedback information probably influences its effectiveness in changing behavior. For example, Eisenberg has found that data feedback only works when it is delivered by clinical leaders.⁵¹ Thus, feedback programs operated by a government regulator or managed care organization may be less effective than professionally based educational programs in which an ongoing relationship exists between the sender and receiver of information.

The level at which feedback is given is another important issue that may differentiate successful and valid programs from questionable ones. For example, many existing DUR programs attempt to review the appropriateness of medication prescribing for individual patients (e.g., drug interactions and dosage). Since the majority of feedback messages are likely to be clinically unimportant,^{14,65} the clinically relevant messages could be unintentionally ignored. For this reason, a more

valid method may be to compare patterns of prescribing by individual physicians with clinical guidelines. Lastly, beyond the medium and the message, if physicians are not able to respond immediately to the feedback delivered, by altering prescribing during a specific patient encounter, they may not respond at all. It is not necessarily true that physicians will generalize behavior from one specific encounter to similar clinical situations.

Unfortunately, given the diversity of feedback approaches, published evaluations of well controlled studies in community populations are rare, and the results are mixed. Computerized feedback on outpatient prescribing charges did not affect prescribing practices in a recent randomized clinical trial.⁶⁶ One noteworthy exception is a randomized clinical trial of computerized feedback by senior physicians to 44 family practitioners in an academic medical center.⁶⁷ Computer messages targeted 28 brand-name drugs with suggested generic alternatives, which were fed back to each physician every month. Each message summarized instances when significant savings could have occurred. After the feedback program began, the rate of generic prescribing in the study group was almost double the rate in the controls. As described above, the ability and influence of the senior physician was probably a key factor in the success of this program, and the results may not be generalizable to nonacademic settings.

REMINDER AND DECISION SUPPORT SYSTEMS

Often, physicians are predisposed to certain therapeutic interventions, but simply omit them due to oversight or lack of coordination in the health care/communications system. In these cases, computerized reminder systems have been developed that enable physicians to reduce these errors of omission by issuing alerts to perform specific actions in response to patient-level information such as laboratory findings or diagnoses.

Several studies in managed care organizations and primary care settings have provided strong evidence that such systems can prevent the omission of essential preventive services such as influenza immunization, hypertension treatment,

and others.⁶⁸⁻⁷⁰ In general, prospective reminders are more effective than retrospective feedback; however, such systems are effective only as long as the reminders continue. Further, it is likely that such systems are only effective when clinicians are already predisposed to acting in concert with the protocols. Few data are available on the potential for such systems to reduce inappropriate drug prescribing in cases when physicians have strong beliefs in opposition to recommended practice. Finally, few well controlled studies are available on the potential for such computerized systems to succeed beyond a "secretarial reminder" function, although early work using advanced decision support systems at Brigham and Women's Hospital in Boston,^{71,72} LDS Hospital in Salt Lake City,^{73,74} and the Regenstrief Institute in Indianapolis^{37,75} shows great promise in altering physicians' prescribing decisions in more complex areas such as dosage, schedule, suboptimal choices, and prevention of adverse drug events.⁷⁰

OPINION LEADERS OR EDUCATIONALLY INFLUENTIAL PHYSICIANS

The role of local opinion leaders in the adoption of new pharmaceutical agents has been well documented by Coleman *et al.*² Their data indicated that after opinion leaders adopted drugs, other less integrated physicians eventually followed in a classic curve of technology diffusion. In several studies of diffusion of scientific information on treatment of arthritis and the inappropriate use of Cesarean sections,^(e.g., 76, 77) local opinion leaders or educationally influential physicians have been identified and encouraged to consult informally with colleagues. These opinion leaders are approached frequently for clinical advice; trusted by their colleagues to evaluate new medical practices in the context of local norms; have good listening skills; and are perceived as clinically competent and caring.^{3, 46, 51} In addition to opinion leader involvement, these interventions generally included brief orientation to research findings, printed educational materials, and encouragement to implement guidelines during informal "teachable moments" that occur naturally in their ongoing collegial associations. Success of these

programs was attributed to "the importance of the local community's norms, the orientation of practitioners to locally credible individuals, and the need to translate the research findings into a locally applicable message."⁷⁸

A recently reported RCT demonstrated that opinion leaders could be used to improve prescribing in the treatment of acute myocardial infarction.⁴⁶ Thirty-seven hospitals in Minnesota were randomized to guideline dissemination, performance feedback, and opinion leaders (intervention) or guidelines and feedback alone (controls). Both the unit of randomization and the unit of analysis were the hospital. Clinical and process data were collected for a year before and a year after the intervention (which itself lasted about 6 months). The opinion leaders were asked to promote four separate practices, each consistent with national evidence-based guidelines: increased use of aspirin, increased use of β -blockers, increased use of thrombolytic therapy in elderly patients, and decreased routine use of lidocaine prophylaxis. Compared to controls, the intervention hospitals successfully increased the use of aspirin (absolute median improvement 13%, $p = 0.04$) and β -blockers (absolute median improvement 31%, $p = 0.02$) (see Figure 30.4). However, there was no improvement in the use of thrombolytic therapy, and all hospitals decreased use of lidocaine by about 50%. This latter finding is evidence of a secular trend, a trend more powerful than the intervention itself, and one that would have been attributed to the intervention if a weaker study design had been employed. Whether or not such interventions can improve prescribing outside the hospital setting, for other conditions, and are cost effective, remains to be determined.

FACE-TO-FACE EDUCATIONAL OUTREACH

A growing number of well controlled studies supports the conclusion that programs combining professionally illustrated educational materials with brief, face-to-face visits (15-25 minutes) by medical-school- or medical-society-based clinical pharmacists (academic detailers) or physician "counselors" are effective in reducing prescribing of contraindicated or marginally effective therapies

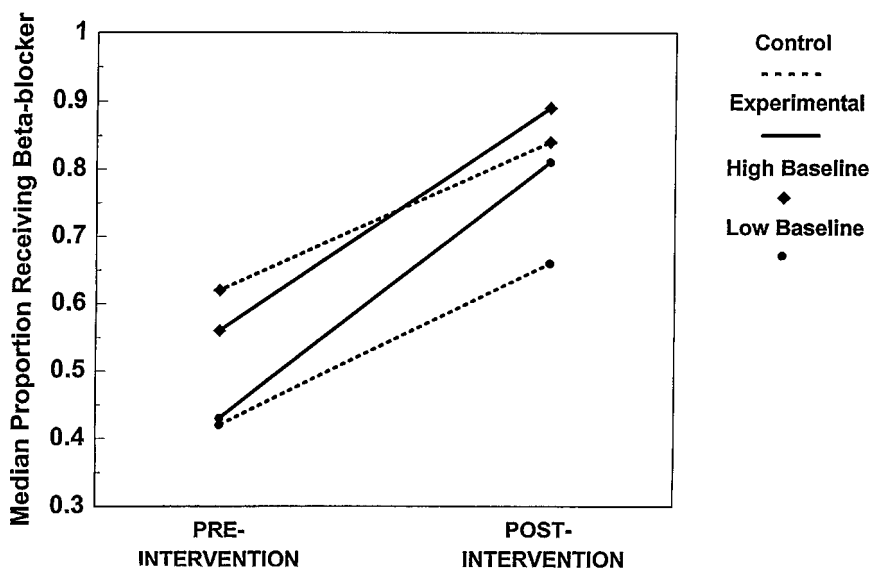


Figure 30.4. Change in proportion of eligible patients receiving β -blockers, stratified by hospitals above and below median baseline proportion. Median proportion was 0.49 (experimental group). The number of hospitals was 11 for the experimental group and seven for the control group. Before and after change represents median change in proportion receiving β -blockers in each of the four strata.⁴⁶

in primary care settings. Similarly, several controlled studies of direct educational efforts by clinical pharmacists have also documented improvements in targeted prescribing practices.⁷⁹ The principles and methods of this approach are described in detail elsewhere¹⁰ and include targeting of physicians with higher than average needs for education (e.g., through analyses of administrative data), conducting motivational research (e.g., surveys of focus group interviews) in advance of the intervention to understand the causes of suboptimal prescribing patterns, sponsorship by authoritative and credible medical organizations, two-way communication with prescribers to increase clinician involvement and relevance to different patient populations and settings, presentation and discussion of counterarguments to which physicians have been opposed, brevity, use of high-quality, graphical educational materials (both for mailings and discussion), repetition of major messages and positive reinforcement, and follow-up reinforcement visits.

Figure 30.5 provides an example of the reverse side of an educational leaflet briefly summarizing the main educational messages concerning the

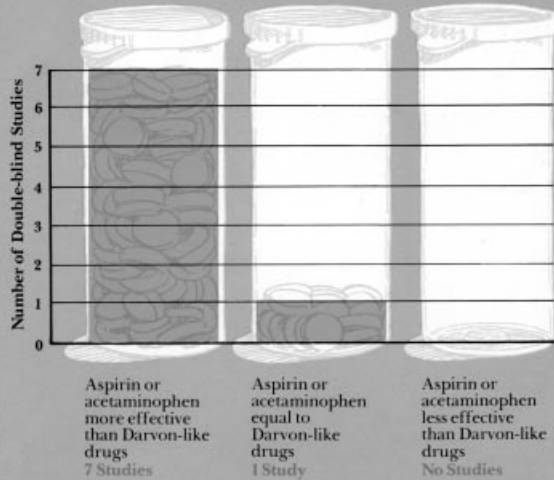
costs and lack of efficacy of propoxyphene that were emphasized in one randomized trial of a four-state academic detailing program. The only formal economic analysis of this approach concluded that targeting moderate to high prescribers of propoxyphene, cephalexin, and peripheral/cerebral vasodilators using administrative claims databases could lead to high benefit-to-cost ratios, even without considering positive spillover effects to nonparticipating physicians, improved quality of care, or possible cost savings due to elimination of adverse drug effects.

If academic detailing is truly cost neutral (or even cost saving), the main barrier to more widespread use of the strategy is its labor intensiveness. Nevertheless, all 15 RCTs of educational outreach identified by the Cochrane Collaboration have demonstrated significant improvements in one or more measures of performance.¹⁹ A number of controlled trials have attempted to replicate the positive results of face-to-face outreach with small group outreach sessions (often referred to as "group detailing"). Group detailing has the additional advantage of encouraging discussions within the group, which may enhance the diffusion of

Efficacy:

Recently, Miller analyzed all double-blind controlled studies of propoxyphene published since 1970.¹ Eight trials compared the drug directly with aspirin or acetaminophen. Here are the results of that review:

Propoxyphene (Darvon, etc.)^{*} versus Aspirin or Acetaminophen²⁻⁹



References

- 1 Miller RR. Propoxyphene: a review. *Drug Therapy Reviews*. Volume 2. Edited by RR Miller, D Greenblatt. New York, Elsevier, 1979, pp 218-257.
- 2 Figure includes all double-blind studies reviewed comparing aspirin or acetaminophen versus propoxyphene in commonly used dosages.
- 3 Sunshine A, Slaughter J, Gruber C: A comparative analgesic study of propoxyphene, fenoprofen, the combination of propoxyphene and fenoprofen, aspirin, and placebo. *Journal of Clinical Pharmacology* 18:556-563, 1978.
- 4 Gruber C, Baptista A, Bauer R, et al: A multicenter analgesic study using single doses of placebo, propoxyphene and acetaminophen. *Journal of Medicine* 8:33-51, 1977.
- 5 Kay B: A clinical comparison of orally administered aspirin, dextropropoxyphene and pentazocine in the treatment of post-operative pain. *Journal of International Medical Research* 2:149-152, 1974.
- 6 Hopkinson JH, Barden FH, Steffens AO, et al: Acetaminophen versus propoxyphene hydrochloride for relief of pain in episiotomy patients. *Journal of Clinical Pharmacology* 13:251-263, 1973.
- 7 Moertel CG, Ahmann DL, Taylor WF, et al: A comparative evaluation of marketed analgesic drugs. *New England Journal of Medicine* 286:813-815, 1972.
- 8 Wang RI, Sandoval RG: The analgesic activity of propoxyphene napsylate with and without aspirin. *Journal of Clinical Pharmacology* 11:310-317, 1971.
- 9 Gruber CM, Baptista A, Kiplinger GF: Relief of postpartum uterine cramping with propoxyphene and aspirin. *Toxicology and Applied Pharmacology* 19:546-553, 1971.
- 10 Based on cost to pharmacist, plus standard one-third mark-up (one-half mark-up for aspirin and acetaminophen). Median costs were used for generic products with multiple manufacturers. Source: *Drug Topics Redbook*. Oradell, New Jersey, Medical Economics, 1980.
- 11 *The Medical Letter* 15:61-62, 1973.
- 12 *AMA Drug Evaluations*. Fourth edition. Chicago, American Medical Association, 1980, pp 75-78.

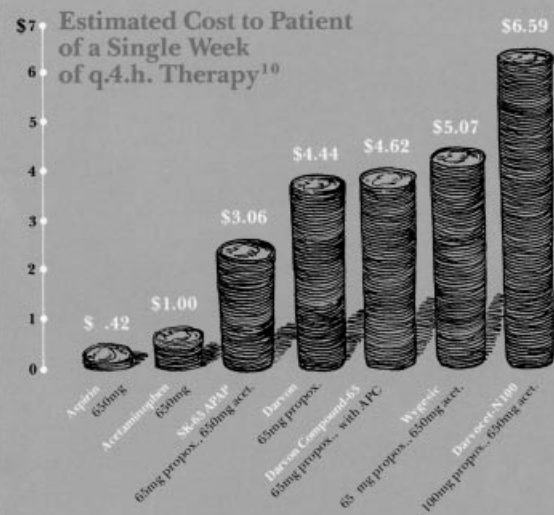
For mild to moderate pain, aspirin or acetaminophen is still the drug of choice.^{11, 12}

Patient resistance to "plain aspirin" can usually be overcome by an explanation of these facts. For the patient who demands a prescription, enteric-coated aspirin offers the safety and efficacy of aspirin without the drawbacks of Darvon-like drugs.

If symptoms are too severe to control with aspirin or acetaminophen, it is unlikely that propoxyphene will control the pain. For acute treatment, codeine may be a better choice.

Cost:

Propoxyphene (Darvon, etc.) offers unimpressive efficacy—at a higher price:



^{*}Propoxyphene and related compounds are marketed under a variety of trade names, including Darvon, Darvon Compound, Darvon-N, Darvocet-N 100, SK-65, S-Pinacet, Unigesic and Wygesic.

This information has been prepared by the Drug Information Program of Harvard Medical School, under the direction of Jerry Avorn, M.D. A brief pamphlet for laypersons, "Why Your Doctor Prescribed 'Plain Aspirin,'" has been written for physicians to distribute to patients in explaining the material presented above. For copies (specify number desired), and for additional information, please write to the Drug Information Program, Harvard Medical School, 643 Huntington Avenue, Boston, MA 02115. The Drug Information Program is supported in part by grant No. HS03880 from the National Center for Health Services Research of the Public Health Service.



Figure 30.5. Reverse side of graphically illustrated and referenced educational leaflet emphasizing the lack of efficacy and high cost of propoxyphene in comparison to aspirin or acetaminophen. The leaflet was used in a face-to-face academic detailing study.^{10, 53}

ideas and increase their impact. For example, in a recent trial to improve the treatment of hyperlipidemia in Sweden, Diwan *et al.* randomized 134 health centers.⁸⁰ The intervention, for 67 of the health centers, consisted of printed guideline dissemination, an informational video, and four 30 min group detailing sessions between all health center physicians and a clinical pharmacist, while the control centers only received printed information. The unit of analysis was the health center. Compared to baseline measurements, hyperlipidemia was treated more often for all patients in the intervention centers, although this reached conventional statistical significance only for women younger than 65 years. Moreover, “first-line prescribing” (specific agents in accordance with national guidelines) was 20% higher ($p = 0.03$) for the intervention centers. It is likely that as long as the group size is relatively small (i.e., fewer than ten participants), and the other precepts of academic detailing are adhered to, group detailing is a reasonable approach to educational outreach. This is an area that merits further research.

FINANCIAL INCENTIVES AND PENALTIES

Although there are few studies of the effects of financial incentives or penalties on physician prescribing behavior *per se*, numerous observational studies suggest that differing payment methods do affect the way that physicians practice medicine.^{6,20,21,51} In addition, it appears that financial incentives are consistently more powerful than penalties in terms of changing behavior.^{6,51} Most methods involve some form of capitation (e.g., capitation-based prescribing budgets). Any savings on prescribing are reinvested in other health services (the model in the UK) or are delivered to physicians as rewards or penalties (the model in the US and Germany).

For example, in response to escalating drug costs, an increasing number of physician organizations in the US are entering into drug risk-sharing arrangements with health maintenance organizations and other managed care plans.⁸¹ These innovative drug payment mechanisms are designed to hold down the cost of drugs by encouraging physicians to prescribe “preferred” drug products

(e.g., generic drugs or those that are under rebate arrangements with manufacturers). Of note, there appears to be only one rigorous, well controlled study examining the effects of capitating physician organizations, and that study did not separate drug information from other services.⁸² This lack of pharmacy data limits our understanding of the dynamics of how and why changing drug payment mechanisms influence the cost and quality of care.

In general, it is believed capitation encourages physicians to examine their prescribing more critically, resulting in the choice of appropriate, effective, and low cost medications. This belief is based on a number of untested assumptions: (i) practices must be large enough to absorb risk, so that costly, but appropriate prescribing decisions for the individual patient are not unduly affected; (ii) performance feedback to prescribers must be timely and provide specific advice about costs, risks, and possible substitutions; and (iii) physicians must understand and be sensitive to differences in drug pricing.²¹ Because it is unlikely that these assumptions (in general) can be met, any intervention using financial incentives must be considered experimental. As such, any strategy employing financial incentives to change prescribing must be studied rigorously, with particular attention to the quality of prescribing, unintended consequences (e.g. cost shifting), and patient outcomes. Reductions in prescribing costs alone are not appropriate or sufficient outcomes. As Bloor and Freemantle have noted, experimental or quasi-experimental studies are needed before sound evidence-based policies can be formulated.²⁰

THE FUTURE

Based on this synthesis of the research literature, it is clear that our knowledge of the characteristics of successful prescribing-improving interventions is growing rapidly. Passive dissemination of drug information or practice guidelines is a necessary but insufficient method for improving most prescribing behaviors. In general, the achievement of long-term changes in practice will depend on inclusion of multiple strategies that predispose, enable, and reinforce desired prescribing behaviors.

The following characteristics recur in successful prescribing-improvement interventions:

- identification of key factors influencing target and alternative prescribing decisions through surveys, focus groups, or in-depth interviews;
- targeting of physicians in need of education (e.g., through review of claims data) and/or local opinion leaders to increase program cost-effectiveness;
- use of credible and objective messengers and materials;
- face-to-face interaction, especially in primary care settings;
- repetition and reinforcement of only a few major messages and behaviors at a time;
- provision of acceptable alternatives to practices to be extinguished;
- brief, graphic educational guidelines to expose and reinforce messages; and
- an emphasis on the goal of improvement in the quality of prescribing, 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. New models are needed to predict the most effective types of intervention for specific problem types. For example, interactive video disc technology is being used to help patients share in decision making with their clinicians by evaluating the risks and benefits of, and preferences for, various treatments of benign prostatic hyperplasia.⁸³ Although this technology might not be cost-effective for everyday prescribing decisions, it may be helpful in situations where particular medications pose both large risks and benefits (e.g., clozapine for psychoses, hormone replacement therapy post-menopause). Are face-to-face interventions (either one on one or in small groups) always necessary to address strongly held incorrect beliefs? Can reminder systems that are so effective in correcting errors of omission change more resistant errors of commission? Lastly, are advanced decision support systems effective, and if so, are they worth the time, effort, and cost necessary to use them?

Practice settings may also influence the choice of interventions to be evaluated. For example, organized systems of clinicians (e.g., medical groups, independent practice associations) may be conducive to participatory approaches in which practicing physicians, and possibly patients, work with a facilitator/educator to explore current practices and barriers to change, and then develop (or modify) practice guidelines along with methods to measure guideline adherence. The group meetings also serve an additional function, as vehicles for active learning. These approaches have been used to upgrade antibiotic use in Mexico⁸⁴ and decrease the use of injectable medications in Indonesia.⁸⁵ Furthermore, many trials of physician behavior change have used trainees or university-affiliated physicians as study subjects. Much more attention needs to be paid to the study of changing the behavior of busy physicians in community practice. Many successful strategies may not be directly transferable from a university hospital to a busy ambulatory clinic.

Most of the studies we reviewed were designed to assess only whether an intervention changed behavior; few studies have undertaken formal cost-benefit analyses. One formal economic analysis of a four-state randomized controlled trial of “academic detailing” found that the intervention actually led to a net savings.⁵⁴ This is a clear illustration of what Eddy described as “getting more for less,” the potential to improve quality and reduce costs simultaneously.⁸⁶ We are unaware of other well controlled studies that compare the costs and benefits of alternative approaches to improving prescribing practice. This is clearly an important research need.

At present, we know that prescribing problems exist, but we know little about their prevalence or determinants. This paucity of data is remarkable considering three-quarters of all physician visits end in the prescription of a drug.¹⁴ In a study of over 30 000 hospital admissions, drug-related complications were common and accounted for 19% of all adverse events.⁸⁷ Bates *et al.* found that almost one-third of adverse drug events occurring in the hospital were preventable.⁸⁸ Less is known about the ambulatory setting; however, the Risk Management Foundation of the Harvard Medical

Institutions reported that about 6% of all professional liability claims were related to medication errors that occurred outside the hospital.⁸⁹ Most of these errors were related to inadequate monitoring, incorrect dosage, improper management of a medication regime, and the use of the wrong drug. These are all errors of commission; the extent of omission (e.g., underuse of effective therapies) has been virtually ignored. A notable exception is a recent study of 5332 elderly survivors of myocardial infarction, that found only 21% of eligible patients received a β -blocker, leading to an estimated 381 excess deaths among the 2952 eligible nonrecipients.⁹⁰

Several factors determine the relative frequency of problems associated with the use of specific medications. Drugs such as the nonsteroidal anti-inflammatory agents, even if prescribed appropriately, are likely to be among the frequently implicated "problem" drugs simply because they are prescribed so commonly. Alternatively, medications such as the benzodiazepines are particularly difficult to monitor due to their variable rate of metabolism across patients or the fact that they are commonly prescribed along with other psychotropic agents. Moreover, many iatrogenic symptoms are nonspecific and difficult to discern from the complex clinical picture of patients with multiple chronic diseases; consequently, drugs commonly prescribed to frail or clinically complex patients may be disproportionately implicated in prescribing problems.

An assessment of the types and rates of problem associated with various drugs could help to differentiate among these various factors. However, previous studies have provided only a gross assessment of the problem. For example, Lesar and colleagues (1990)⁹¹ reported that of the problems they identified through chart review, antimicrobial agents were the most common drug class involved in prescribing problem errors (23% of the problems), and cardiovascular drugs were the next most common (10% of all problems). However, it is unclear whether this was due to the disproportionate rate of prescriptions for these drugs, an unusually high likelihood of all problems for each prescription, or both factors. Future research efforts need to describe in greater detail

the nature, prevalence, rate of prescribing, and severity of prescribing problems associated with the overuse, underuse, and misuse of medications.

Further, we lack information regarding what proportion of an outpatient population "at risk" (i.e., on targeted problematic drugs) subsequently suffers adverse outcomes such as increased severity of illness, diminished quality of life, drug-related use of medical care services, and mortality. Needed, for example, is a more precise estimate of drug-related problems related to increased medical service use and an analysis of the proportion of these that may be preventable.

Characterization of the multiple and complex factors contributing to adverse drug-related patient outcomes will help to promote safe and effective drug use by alerting physicians to drug-related problems that can be treated immediately. Further, once we identify specific physician prescribing practices and patient drug use behaviors that are related to adverse health outcomes, we can target appropriate interventions designed to reduce the incidence of these outcomes.

Finally, linkage between intervention-induced improvements in prescribing practice and changes in patient outcomes would be an enormous accomplishment in this nascent field. While regulatory-induced reductions in use of essential medications have been associated with increased institutionalization among frail elderly³⁶ and partial hospitalizations among schizophrenic patients,²³ very few analogous patient outcome studies exist in the literature on prescribing education. Because of the important effects of medications on many health outcomes that have been demonstrated in clinical trials, it is reasonable to hypothesize that more appropriate use of some medications could reduce morbidity and mortality, increase patient functioning, and improve quality of life. Whether improved prescribing is a surrogate measure, or an outcome that directly leads to improved health outcomes, it remains a critically important area for study in the next decade and beyond. Furthermore, the promise of a comprehensive electronic patient record, one that is knowledge generating and linked to prescriptions, clinical and laboratory information, and claims data, has yet to be realized.⁹² When it does come to

pass, the fields of health services research and pharmacoepidemiology will enter a new era when innovative measures to improve the quality of prescribing will be studied with heretofore unknown methodologic rigor.

REFERENCES

- Morgan PP. Are physicians learning what they read in journals? *Can Med Assoc J* 1985; **133**: 263.
- Coleman J, Katz E, Menzel H. *Medical Innovation: a Diffusion Study*. Indianapolis, IN: Bobbs-Merrill, 1966.
- Rogers E. *Diffusion of Innovations*. New York: Free Press, 1995.
- Kotler P, Roberto E. *Social Marketing. Strategies for Changing Public Behaviour*. New York: Free Press, 1989.
- Fox R, Mazmanian P, Putnam RW. *Changing and Learning in the Lives of Physicians*. New York: Praeger, 1989.
- Greco PJ, Eisenberg JM. Changing physicians' practices. *New Engl J Med* 1993; **329**: 1271-4.
- 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.
- Gurwitz J, Soumerai SB, Avorn J. Improving medication prescribing and utilization in the nursing home. *J Am Geriatr Soc* 1990; **38**: 542-52.
- Lipton HL, Bird JA. Drug utilization review in ambulatory settings: state of the science and directions for outcomes research. *Med Care* 1993; **31**: 1069-82.
- Soumerai SB, Avorn JL. Principles of educational outreach ("academic detailing") to improve clinical decisionmaking. *J Am Med Assoc* 1990; **263**: 549-56.
- Soumerai SB. Factors influencing prescribing. *Aust J Hosp Pharm* 1988; **18**: 9-16.
- Lipton HL, Lee PR. *Drugs and the Elderly: Clinical, Social and Policy Perspectives*. Stanford, CA: Stanford University Press, 1988.
- 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.
- Soumerai SB, Lipton HL. Computer based drug utilization review—risk, benefit, or boondoggle? *New Engl J Med* 1995; **332**: 1641-4.
- Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *Can Med Assoc J* 1995; **153**: 1423-31.
- Balas EA, Boren SA, Brown GD, *et al*. Effect of physician profiling on utilization: meta-analysis of randomized clinical trials. *J Gen Intern Med* 1996; **11**: 584-90.
- Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; **342**: 1317-22.
- Robinson MB. Evaluation of medical audit. *J Epidemiol Community Health* 1994; **48**: 435-40.
- Bero L, Grilli R, Grimshaw J, Oxman A, eds. *Collaboration on Effective Professional Practice Module of the Cochrane Database of Systematic Reviews*, Issue 1, The Cochrane collaboration. Oxford: Update Software, 1998.
- Bloor K, Freemantle N. Lessons from international experience in controlling pharmaceutical expenditure II: influencing doctors. *Br Med J* 1996; **312**: 1525-7.
- Soumerai SB, Ross-Degnan D. Prescribing budgets: economic, clinical, and ethical perspectives. *Aust Prescriber* 1997; **20**: 28-9.
- Soumerai SB, Ross-Degnan D, Fortess EE, Abelson J. A critical analysis of studies of state drug reimbursement policies: research in need of discipline. *Milbank Q* 1993; **71**: 217-52.
- Soumerai SB, McLaughlin TJ, Ross-Degnan D, Casteris CS, Bollini P. Effects of limiting Medicaid drug reimbursement benefits on the use of psychotropic agents and acute mental health services by patients with schizophrenia. *New Engl J Med* 1994; **331**: 650-6.
- Smalley WE, Griffin MR, Fought RL, Sullivan L, Ray WA. Effect of a prior authorization requirement on the use of nonsteroidal anti-inflammatory drugs by Medicaid patients. *New Engl J Med* 1995; **332**: 1612-7.
- Ray WA, Griffin MR, Shorr RI. Adverse drug reactions and the elderly [commentary]. *Health Aff* 1990; **9**: 114-22.
- Stason WB, Weinstein MC. Allocation of resources to manage hypertension. *New Engl J Med* 1977; **296**: 732-9.
- Soumerai SB, Avorn J, Gortmaker S, Hawley S. Effect of government and commercial warnings on reducing prescription misuse: the case of propoxyphene. *Am J Public Health* 1987; **77**: 1518-23.
- Ross-Degnan D, Soumerai SB, Fortess EE, Gurwitz J. Examining product risk in context: market withdrawal of zomepirac sodium as a case study. *J Am Med Assoc* 1993; **270**: 1937-42.
- Kunin CM. Resistance to antimicrobial drugs—a worldwide calamity. *Ann Intern Med* 1993; **118**: 557-61.
- Beers M, Avorn J, Soumerai SB, Everitt DE, Sherman DS, Salem S. Psychoactive medication use in intermediate-care facility residents. *J Am Med Assoc* 1988; **260**: 3016-20.

31. Avorn J, Soumerai SB, Everitt DE, Ross-Degnan D, Beers MH, Sherman D, Salem-Schatz SR, Fields D. A randomized trial of a program to reduce the use of psychoactive drugs in nursing homes. *New Engl J Med* 1992; **327**: 168–73.
32. Barnett GO, Winickoff RN, Morgan MM, Zielstorff RD. A computer-based monitoring system for follow-up of elevated blood pressure. *Med Care* 1983; **21**: 400–9.
33. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; **89**: 630–8.
34. Donovan M, Dillon P, McGuire L. Incidence and characteristics of pain in a sample of medical–surgical patients. *Pain* 1987; **30**: 69–78.
35. Soumerai SB, Avorn J, Ross-Degnan D, Gortmaker S. Payment restrictions for prescription drugs in Medicaid: Effects on therapy, cost, and equity. *New Engl J Med* 1987; **317**: 550–6.
36. Soumerai SB, Ross-Degnan D, Avorn J, McLaughlin TJ, Choodnovskiy I. Effects of Medicaid drug-payment limits on admission to hospitals and nursing homes. *New Engl J Med* 1991; **325**: 1072–7.
37. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectibility of man. *New Engl J Med* 1976; **295**: 1351–5.
38. Schwartz RK, Soumerai SB, Avorn J. Physician motivations for non-scientific drug prescribing. *Soc Sci Med* 1989; **28**: 577–82.
39. 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.
40. Gilbert JR, Light RJ, Mosteller F. Assessing social innovations: an empirical base for policy. In: Bennett CH, Lumsdaine AA, eds, *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*. New York: Academic, 1975.
41. Cook TD, Campbell DT. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin, 1979.
42. Soumerai SB, Ross-Degnan D, Gortmaker S, Avorn J. Withdrawing payment for non-scientific drug therapy: intended and unexpected effects of a large-scale natural experiment. *J Am Med Assoc* 1990; **263**: 831–9.
43. Groves R. Therapeutic drug-use review for the Florida Medicaid Program. *Am J Hosp Pharm* 1985; **42**: 316–9.
44. Divine GW, Brown JT, Frazier LM. The unit of analysis error in studies about physicians’ patient care behavior. *J Gen Intern Med* 1992; **7**: 623–9.
45. Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford: Clarendon, 1996.
46. Soumerai SB, McLaughlin TJ, Gurwitz JH, et al. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *J Am Med Assoc* 1998; **279**: 1358–63.
47. Avorn J, Soumerai SB, Wessels M, Taylor W, Janousek J, Weiner M. Reduction of incorrect antibiotic dosing through a structured educational order form. *Arch Intern Med* 1988; **148**: 1720–4.
48. Marciniak TA, Ellerbeck EF, Radford MJ, et al. Improving the quality of care for Medicare patients with acute myocardial infarction: results from the Cooperative Cardiovascular Project. *J Am Med Assoc* 1998; **279**: 1351–7.
49. Brook RH, McGlynn EA. Measuring quality of care. *New Engl J Med* 1996; **335**: 966–9.
50. Green LW, Kreuter MW. *Health Promotion Planning: an Educational and Environmental Approach*. Mountain View, CA: Kaiser Foundation, 1991.
51. Eisenberg JM. *Doctors’ Decisions and the Cost of Medical Care*. Ann Arbor, MI: Health Administration Press Perspectives, 1986.
52. Schaffner W, Ray WA, Federspiel CF, Miller WO. Improving antibiotic prescribing in office practice: a controlled trial of three educational methods. *J Am Med Assoc* 1983; **250**: 1728–32.
53. Avorn JL, Soumerai SB. Improving drug therapy decisions through educational outreach: A randomized controlled trial of academically-based “detailing.” *New Engl J Med* 1983; **308**: 1457–63.
54. Soumerai SB, Avorn JL. Economic and policy analysis of university-based drug “detailing”. *Med Care* 1986; **24**: 313–31.
55. Kosecoff J, Kanouse DE, Rogers WH, et al. Effects of the National Institutes of Health consensus development program on physician practice. *J Am Med Assoc* 1987; **258**: 2708–13.
56. Lomas J, Anderson GM, Domnick KP, et al. Do practice guidelines guide practice? The effect of consensus statements on the practice of physicians. *New Engl J Med* 1989; **321**: 1306–11.
57. Wade OL, Hood H. Prescribing of drugs reported to cause adverse reactions. *Br J Prev Soc Med* 1972; **26**: 205–11.
58. Maclure M, Dormuth C, Naumann T, et al. Influences of educational interventions and adverse news about calcium channel blockers on first line prescribing of antihypertensive drugs to elderly people in British Columbia. *Lancet* 1998; **352**: 943–8.
59. Soumerai SB, Ross-Degnan D, Spira J. The effects of professional and media warnings about the association between aspirin use in children and Reye’s Syndrome. *Milbank Q* 1992; **70**: 155–82.
60. Inui TS, Yourtee EL, Williamson JW. Improved outcomes in hypertension after physician tutorials: a controlled trial. *Ann Intern Med* 1976; **84**: 646–51.
61. Klein LE, Charache P, Johannes RS. Effect of physician tutorials on prescribing patterns of graduate physicians. *J Med Educ* 1981; **56**: 504–11.

62. Johnson RE, Campbell WH, Azevedo D, Christensen DB. Studying the impact of patient drug profiles in an HMO. *Med Care* 1976; **14**: 799–807.
63. Hershey CO, Porter DK, Breslau D, Cohen DI. Influence of simple computerized feedback on prescription changes in an ambulatory clinic: a randomized clinical trial. *Med Care* 1986; **24**: 472–81.
64. Koepsell TD, Gurtel A, Diehr PH, *et al.* The Seattle evaluation of computerized drug profiles: effects on prescribing practices and resource use. *Am J Public Health* 1983; **73**: 850–5.
65. Lipton HL, Bero LA, Bird JA, McPhee SJ. The impact of clinical pharmacists' consultations on physicians' geriatric drug prescribing: a randomized controlled trial. *Med Care* 1992; **30**: 646–58.
66. Hershey CO, Goldberg HI, Cohen DI. The effect of computerized feedback coupled with a newsletter upon outpatient prescribing charges. A randomized controlled trial. *Med Care* 1988; **26**: 88–93.
67. Gehlbach SH, Wilkinson WE, Hammond WE, *et al.* Improving drug prescribing in a primary care practice. *Med Care* 1984; **22**: 193–201.
68. Barnett GO, Winickoff RN, Dorsey JL, *et al.* Quality assurance through automated monitoring and concurrent feedback using a computer-based medical information system. *Med Care* 1978; **16**: 962–70.
69. Tierney WM, Hui SL, McDonald C. Delayed feedback of physician performance versus immediate reminders to perform preventive care: effects of physician compliance. *Med Care* 1986; **24**: 659–66.
70. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *J Am Med Assoc* 1998; **280**: 1339–46.
71. Bates DW, Kuperman G, Teich JM. Computerized physician order entry and quality of care. *Qual Manage Health Care* 1994; **2**: 18–27.
72. Bates DW, Leape LL, Cullen DJ, *et al.* Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *J Am Med Assoc* 1998; **280**: 1311–16.
73. Evans RS, Pestotnik SL, Classen DC, *et al.* A computer assisted management program for antibiotics and other anti-infective agents. *New Engl J Med* 1998; **338**: 232–8.
74. Majumdar SR, Ross-Degnan R, Soumerai SB. A computer-assisted management program for antibiotics and other anti-infective agents (comment). *New Engl J Med* 1998; **338**: 775.
75. McDonald CJ, Hui SL, Smith DM, *et al.* Reminders to physicians from an introspective computer medical record: a two-year randomized trial. *Ann Intern Med* 1984; **100**: 130–8.
76. Stross JK, Bole GG. Evaluation of a continuing education program in rheumatoid arthritis. *Arthritis Rheum* 1980; **23**: 846–9.
77. Lomas J, Enkin M, Anderson G, *et al.* Opinion leaders vs. audit and feedback to implement practice guidelines. Delivery after previous cesarean section. *J Am Med Assoc* 1991; **265**: 2202–7.
78. Lomas J. Teaching old (and not so old) dogs new tricks: implementing research findings in primary care. In: Lomas J, ed., *Disseminating New Knowledge and Having an Impact on Practice*. New York: Sage, 1994.
79. Stergachis A, Fors M, Wagner EH, *et al.* Effect of clinical pharmacists on drug prescribing in a primary care clinic. *Am J Hosp Pharm* 1987; **44**: 525–9.
80. Diwan VK, Wahlstrom R, Tomson G, *et al.* Effects of group detailing on the prescribing of lipid lowering drugs: a randomized controlled trial in Swedish primary care. *J Clin Epidemiol* 1995; **48**: 705–11.
81. Lipton HL, Kreling DK, Collins T, Hertz KC. Pharmacy benefit management companies: dimensions of performance. *Annu Rev Public Health* 1999; **20**: 361–401.
82. Robinson JC, Casalino LP. The growth of medical groups paid through capitation in California. *New Engl J Med* 1995; **333**: 1684–7.
83. Barry MJ, Cherkin DC, Chang Y, Fowler FJ, Skates S. A randomized trial of a multi-media shared decision-making program for men facing a treatment decision for benign prostatic hyperplasia. *Dis Manage Clin Outcomes* 1997; **1**: 5–14.
84. Gutierrez G, Guiscafre H, Munoz O. Strategies for improving the therapeutic patterns used in acute diarrhea in primary medical care units X. Conclusions and research perspectives. *Arch Invest Med* 1988; **19**: 437–43.
85. Hadiyono JEP, Suryawati S, Danu SS, Sunartono S, Santosa B. Interactional group discussion: results of a controlled trial using a behavioral intervention to reduce the use of injections in public health facilities. *Soc Sci Med* 1996; **42**: 1177–83.
86. Eddy D. Rationing resources while improving quality: how to get more for less. *J Am Med Assoc* 1994; **272**: 817–24.
87. Leape LL, Brennan TA, Laird NA, *et al.* The nature of adverse events in hospitalized patients. *New Engl J Med* 1991; **324**: 377–83.
88. Bates DW, Cullen DJ, Laird N, *et al.* Incidence of adverse drug events and potential adverse drug events: implications for prevention. *J Am Med Assoc* 1995; **274**: 29–34.
89. Dwyer K. *Medication Related Malpractice Claims*. Forum, Risk Management Foundation of the Harvard Medical Institutions, 1998; 4–8.
90. Soumerai SB, McLaughlin TJ, Spiegelman D, *et al.* Adverse outcomes of underuse of beta blockers in

- elderly survivors of acute myocardial infarction. *J Am Med Assoc* 1997; **277**: 115–21.
91. Lesar TS, Briceland LL, Delcours K, Parmalee JC, Masta-Gornic V, Pohl H. Medication prescribing errors in a teaching hospital. *J Am Med Assoc* 1990; **263**: 2329–34.
92. Schiff GD, Rucker TD. Computerized prescribing: building the infrastructure for better medication usage. *J Am Med Assoc* 1998; **279**: 1024–9.

31

Drug Utilization Review

SEAN HENNESSY

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

BRIAN L. STROM

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

HELENE L. LIPTON

Department of Clinical Pharmacy, University of California at San Francisco Schools of Pharmacy and Medicine, USA

STEPHEN B. SOUMERAI

Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA, USA

INTRODUCTION

Drug utilization review (DUR) is one of a growing number of approaches to improve quality and reduce costs in health care. As the term implies, DUR programs focus on medication use problems, which have been well described.^{1,2} 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.”³ In this chapter, we distinguish DUR programs from similar activities, describe the conceptual framework that underlies DUR, review the history of DUR, and describe the DUR process and how it is applied in different settings. We then describe the clinical and methodologic problems to be ad-

ressed by pharmacoepidemiology research in this area, critically evaluate the evidence regarding the effectiveness of DUR programs, and look to DUR’s future. While DUR programs hold promise for improving the quality of drug therapy, there is currently no credible evidence that DUR programs improve patient outcomes. Continued support of DUR programs should be conditional on credible evidence that they are effective.

DUR has many synonyms, including *drug use review*, *drug use evaluation*, and *medication use evaluation*. DUR programs differ from *drug utilization studies* (see Chapter 29), which are time limited investigations that quantify drug use, but do not assess appropriateness or attempt to change practice. Although *DUR programs* and *DUR studies* both measure and assess the appropriateness of

therapy, DUR programs are ongoing and include an intervention component, while DUR studies are time limited and do not intervene to change therapy.

DUR programs operate in many different settings, and within the context of other programs that seek to improve quality and reduce costs in health care. One such program is the formulary system, which makes use of *formularies*, or enumerated lists of drugs that may be prescribed (and, through exclusion, those that may not be prescribed). Other administrative programs that affect medication use within drug benefit programs include the imposition of limits (so called “caps”) on the monetary cost of prescription medications or on the number of prescriptions that will be paid per person in a given time period, required co-payments by consumers, mandatory generic drug substitution, and requirements for prior approval for particular drugs. DUR is also related to monthly, federally mandated drug regimen reviews performed by pharmacists in US nursing homes. Finally, DUR is intertwined with contemporary pharmacy practice, which involves reviewing patient-specific information maintained by the individual pharmacy (i.e., the patient profile) each time a prescription is filled.⁴ As discussed later, some programs calling themselves DUR rely exclusively on such reviews.

DUR programs are often categorized by the timing of interventions within the drug use process. Unfortunately, the use of such terms has been inconsistent. In this chapter, we will classify DUR programs as *prospective* if they intervene before the patient receives the medication and *retrospective* if they intervene after the patient has received the medication.

CONCEPTUAL FRAMEWORK UNDERLYING DUR

Lipton and Bird⁵ have elucidated a conceptual framework underlying DUR, which is presented schematically in Figure 31.1. According to this model, prospective and retrospective DUR programs affect either current or future prescribing and dispensing practices within the context of other important factors. These factors can be categorized as patient and family influences, prescriber characteristics, and system factors (Figure 31.1). Patient

and family influences include demographic characteristics, cultural beliefs, unwillingness to take needed drugs, and demands for unnecessary drugs (e.g., demands for antibiotics in the setting of viral infections, which do not respond to antibiotics). Prescriber characteristics include inadequate knowledge about pharmacology, inadequate knowledge of drug prices, and forgetfulness. System factors include drug reimbursement policies, formularies, organization of the medical care system, influence of pharmaceutical companies, abundance of drug therapy options, and lack of data on comparative safety and efficacy among therapeutic alternatives.

DUR programs are theorized to counteract many of the detrimental influences of patient, prescriber, and system factors. Also reflected in this model is the fact that both prospective and retrospective DUR might influence care for patients identified in the alerts, the care of future patients (called the “spillover effect”), or both. Finally, integration of prospective and retrospective DUR might hold greater promise for improving prescribing than would their use in an uncoordinated fashion. Naturally, this possibility deserves to be evaluated empirically.

HISTORY OF DUR

The literature on DUR programs, and particularly on outpatient DUR programs, has emanated predominantly from North America. Several developments are viewed as having sentinel importance in the emergence of DUR in the US.⁵⁻⁹ One such factor is the growth of insurance coverage for outpatient prescriptions that took place in the 1970s and 1980s,⁷ which created both the financial interest to minimize prescription drug costs and the computerized patient-level data needed to conduct large scale programs. Technical feasibility within US Medicaid (the federally sponsored health care benefit program for the poor) was enhanced in the early to mid-1960s, when the US Department of Health, Education and Welfare (DHEW, predecessor of the present US Department of Health and Human Services (DHHS)) collaborated with state Medicaid agencies to develop computerized information systems to perform administrative functions within Medicaid. The first published reference to

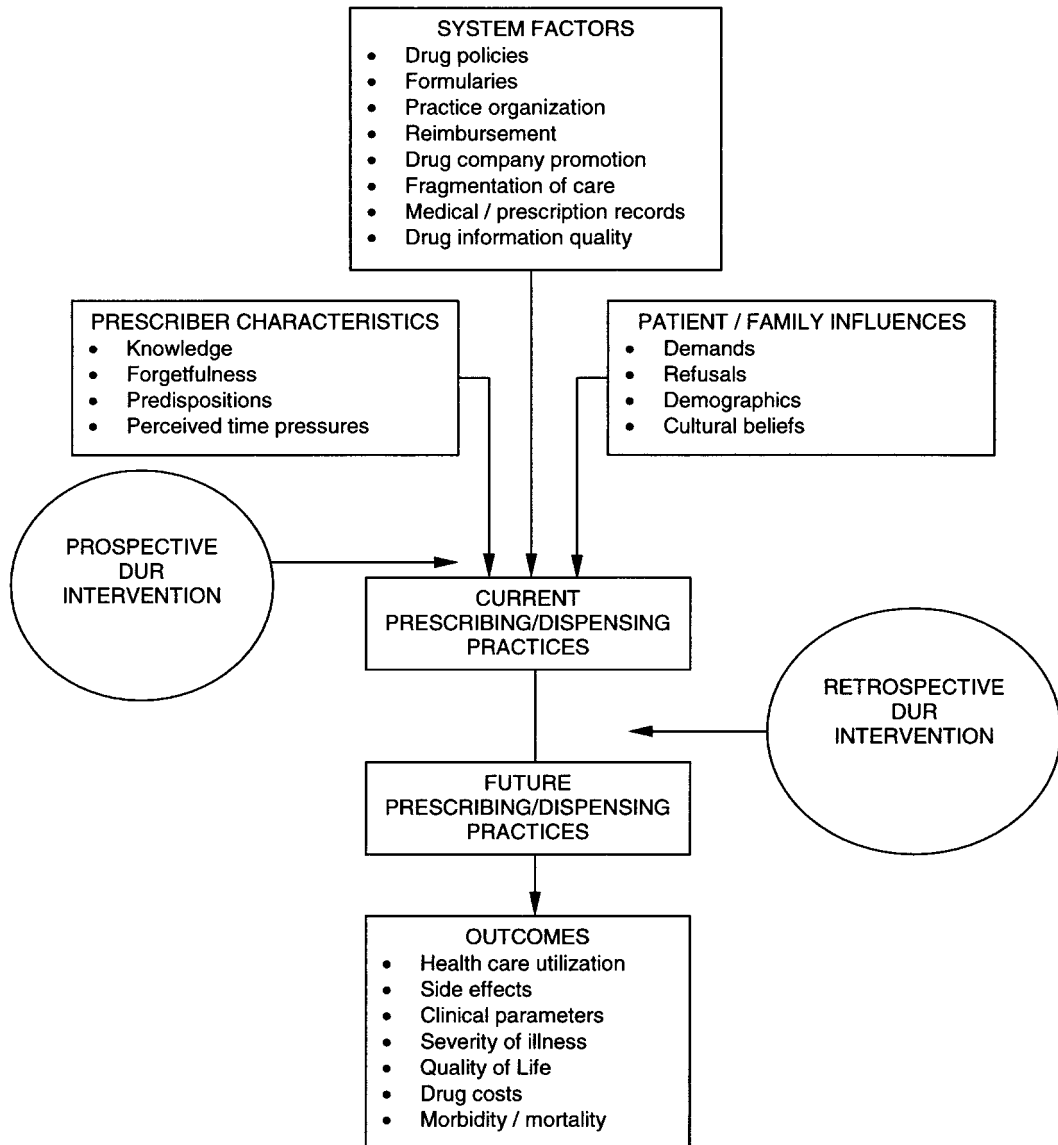


Figure 31.1. Relationship of DUR interventions to drug prescribing and dispensing. Reprinted with permission from Lipton HL, Bird JA. Drug utilization review in ambulatory settings: state of the science and directions for outcomes research. *Med Care* 1993; 31: 1069–82.

DUR was a 1969 background paper by the US DHEW’s Task Force on Prescription Drugs, which was charged with evaluating the feasibility and implications of providing prescription drug coverage to beneficiaries of the US Medicare system. As part of this evaluation, the Task Force raised the possibility of examining claims data in order to improve the quality and reduce the cost of drug

therapy.¹⁰ Although the Task Force perceived that outpatient DUR appeared promising,¹⁰ it believed that evidence of effectiveness was needed before implementation, and recommended further study of DUR rather than its widespread adoption.¹¹ However, this did not occur.

In 1970, PAID Prescriptions, a private sector pharmaceutical benefit management company in

the US, began a formal DUR program in California's San Joaquin Valley Medicaid program that focused on cost issues.^{12,13} By the mid-1970s, other US Medicaid agencies and other third party payers had begun collaborating with private companies, known as DUR vendors, to initiate outpatient DUR programs that focused both on cost and quality of care issues. Based largely on the intuitive appeal and promise of DUR, it was enthusiastically advocated in a 1977 report that examined issues surrounding a drug benefit for a proposed US national health insurance program.¹⁴ The Joint Commission on Accreditation of Healthcare Organizations (JCAHO), which accredits health care facilities in the US, required hospitals to conduct inpatient DUR for antibiotics beginning in 1985, and for medications in general beginning in 1987 (Rich D, personal communication, 13 October 1998). By that time, DUR had been practiced in some hospitals for over a decade.¹⁵⁻¹⁸ In 1988, DUR programs were required to be part of a US Medicare drug benefit,¹⁹ although this mandate was revoked when the legislation providing the benefit was repealed the following year.²⁰ Probably the most significant event in the growth of outpatient DUR in the US was the Omnibus Budget Reconciliation Act of 1990,²¹ which mandated that all US states conduct both retrospective and prospective DUR on behalf of Medicaid enrollees.

A 1994 report by Walser⁹ indicated that outpatient DUR programs were being adopted more frequently outside of the US, as well. However, literature describing and evaluating such programs is lacking.

THE DUR PROCESS

Although there is some variation in how various authors specify the steps of the DUR process,^{6,13,22,23} one general and widely accepted model for both retrospective and prospective DUR was described by Erwin:²²

1. Design the basic structure.
2. Seek approval.
3. Construct criteria.
4. Apply criteria.
5. Evaluate and analyze yield.
6. Establish prescribing patterns.
7. Establish intervention strategies.
8. Re-apply criteria to databases.
9. Revise criteria as needed.

Although this basic model is applied differently in different settings, some characteristics are common across settings. In general, 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, apparent abuse of psychoactive medications, and use of a more costly agent when a less costly agent is available. However, it has been observed that DUR programs tend to focus on errors of commission, giving less attention to problems of underuse.²⁴

Naturally, a criterion must be *valid* in order for the application of that criterion to improve clinical outcomes. In this context, we define validity as the ability of a DUR criterion to detect instances in which a change in therapy would ultimately benefit the patient. However, because no outpatient DUR program has ever been shown to reduce morbidity or mortality, the validity of the criteria used in outpatient DUR programs is unknown. Further, there is good reason to withhold judgement on the validity of any DUR criterion until it has been demonstrated empirically. Specifically, the current state of knowledge in drug therapy is such that what is considered to be "good care" is supported by levels of evidence that range from case reports, to reasoning based in pharmacologic theory, to conclusive randomized trials. Even when DUR criteria are based on the results of conclusive randomized trials (i.e., the best case scenario), it remains uncertain that the criteria will identify patients whose care will improve if therapy is

changed. For example, the use of β -blockers has been conclusively shown to improve survival in patients who have had a heart attack.²⁵⁻²⁷ Thus, a very reasonable criterion would be that patients with a history of myocardial infarction and no evidence of a clinical condition that would preclude the use of a β -blocker (e.g., chronic obstructive lung disease) should receive a β -blocker. However, it is conceivable that applying this criterion in real life would identify patients who, because of some measured or unmeasured characteristics, would fail to benefit from β -blocker therapy. This is because patients in whom criteria are violated may well be different from the population in whom β -blockers were tested. Thus, the only way to know for certain whether a given criterion is valid is to measure the clinical effectiveness of a DUR program that employs that criterion.

The next step in the DUR process is to measure adherence to 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. 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 exception. Naturally, the intervention needs to be successful in achieving a *beneficial* change in therapy (rather than *any* change in therapy, or even the *desired* change in therapy) for a DUR program to be effective. The effectiveness of different approaches to changing prescribing is discussed in detail in Chapter 30.

General differences in how the general DUR model is applied in different settings will now be discussed.

OUTPATIENT RETROSPECTIVE DUR PROGRAMS

Although US federal law requires outpatient retrospective DUR programs only for Medicaid patients, most non-Medicaid outpatient drug

benefit plans in the US also perform retrospective DUR. Because of the large number of patients, virtually all outpatient retrospective DUR programs use computerized administrative data (i.e., data maintained for billing and other administrative purposes) to identify exceptions. Some DUR programs use both pharmacy and medical claims data, while others use pharmacy claims only. Naturally, one would expect that programs that use medical claims data (which typically include diagnoses) would be in a better position to improve care than those that do not have access to diagnoses. However, this expectation deserves to be studied empirically.

Although the frequency of exceptions varies with the criteria employed, one study showed that, each year, 3-7% of persons who filled a prescription had an exception to a set of 61 DUR criteria.²⁸ In most outpatient retrospective DUR programs, computer generated exceptions are reviewed by a physician or pharmacist, or by a committee of health professionals, and result in an intervention only when the exception does not meet some implicit or subjective criterion.

Most outpatient retrospective DUR programs intervene by mailing an alert letter to the physician, and sometimes to the pharmacist of record. This 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. A few outpatient retrospective DUR programs use telephone calls or face-to-face meetings to convey alerts. An ongoing development in outpatient retrospective DUR is the use of data grouped at the level of the prescriber (an approach known as "physician profiling") to make decisions regarding interventions, rather than making intervention decisions by viewing each exception in isolation.^{29,30} This approach has intuitive appeal, and its potential advantages deserve to be studied.

Recent years have also seen a dramatic rise in the proportion of US Medicaid recipients who are enrolled in managed care (rather than fee-for-service) programs, with an increase from 12 to 40% of the covered population from 1992 to 1996 (source, Health Care Financing Administration world wide web site: <http://www.hcfa.gov/medicaid>).

Because managed care companies are often responsible for performing DUR within their enrollee population, DUR within Medicaid appears to be becoming increasingly decentralized. As discussed below, this decentralization may hamper efforts to conduct rigorous studies of the effectiveness of DUR.

OUTPATIENT PROSPECTIVE DUR PROGRAMS

Prospective DUR includes efforts that occur before or during the prescribing process, or during the dispensing process. Thus, the use of computerized decision support programs, whether by physicians³¹ or by pharmacists,³²⁻³⁵ is a form of prospective DUR. Computer applications have been developed for use by physicians in outpatient settings for managing particular therapies such as oral anticoagulants,^{36,37} and to reduce drug costs.³⁸⁻⁴² However, in the US at least, computer entry of outpatient prescriptions by physicians remains rare.³¹

In current US practice, most outpatient prospective DUR programs intervene by conveying computer alerts to the pharmacist filling the prescription. If the pharmacist is able to override the alert and dispense the medication, the alert is known as a "soft edit." If the DUR program refuses payment for the medication in question, it is called a "hard edit." The responsibility of resolving the potential problem then 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 (who is in a better position to change therapy) would be expected to hamper prospective DUR's effectiveness in changing drug therapy.

Manual review by the pharmacist of patient profiles is a form of prospective DUR, and has been a standard component of pharmacy practice for decades.⁴ Computer aided review of patient profiles, also a form of prospective DUR, has become increasingly common in recent years. Computer aided review may have advantages over manual review because of increased sensitivity in detecting potential problems. However, it is also

possible for systems to miss important problems while providing a false sense of security to the user. As well, computer review systems that frequently produce apparently trivial alerts may foster disregard of alerts, reducing the potential effectiveness of such systems.

A distinction can be drawn between *on-line* prospective DUR, which uses a centralized profile containing all of a patient's prescriptions, regardless of which pharmacies have filled them, and *in-pharmacy* prospective DUR, which uses data from a single pharmacy or pharmacy chain. Centralized systems would be expected to have an advantage over in-pharmacy systems because of the former's more complete patient-specific database. This potential advantage deserves to be studied. In addition, some US states fulfill their federal mandate to perform prospective DUR for Medicaid patients by requiring pharmacists in their state to perform in-pharmacy prospective DUR, whether by manual or computer aided methods.

Currently available on-line prospective DUR programs do not use diagnosis data, which would be expected to limit the ability of prospective DUR programs to identify drug-disease interactions and omitted-but-necessary therapies. Finally, unlike retrospective DUR, prospective DUR incorporates no human reviewer to serve as a filter between raw computer output and the person receiving the alert. If a large proportion of such alerts is perceived to be falsely positive, the alerts may become increasingly ignored, thus limiting the effectiveness of prospective DUR programs.

HOSPITAL DUR PROGRAMS

Hospitalized patients are sicker, and take more numerous and toxic medications than community dwelling individuals. Hospitals are also characterized by more centralized decision making and better availability of clinical data than is present in most outpatient environments. Therefore, it is not surprising that DUR originated in hospitals, where it is both more acutely needed and more politically and logistically feasible.

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. They use the same overall process that is described above, although they generally do so on a more limited number of patients than do most outpatient programs. As a result, many hospital DUR programs obtain data through manual techniques such as chart review in addition to using computerized data. Hospital DUR programs tend to perform series of discrete evaluations rather than ongoing evaluations, and often include elements of both prospective and retrospective review.

A growing number of hospitals use computer tools to evaluate therapy as physicians enter *drug orders* (as prescriptions are called in hospitals) into a computer. However, even the direct physician order entry system described by Bates *et al.*,⁴³ which is advanced by today's standards, was designed to reduce dosing and transcription errors, rather than to improve drug therapy by identifying drug–drug interactions, drug–disease interactions, and the other types of medication problem usually targeted by DUR programs.

Because of more centralization in decision making authority, hospital DUR programs are often able to use regulatory interventions that are unavailable to many outpatient DUR programs. These interventions can include formulary deletions, restriction of a medication for use in particular circumstances, restriction of a medication for use by certain specialists, and mandatory drug order forms, which tend to be more effective than nonregulatory interventions in affecting prescribing.⁴⁴

Therapeutic areas that have received recent attention by DUR programs include antimicrobials,^{45–73} analgesics,^{74–80} sedative–hypnotics,^{74, 75, 79–83} and anti-ulcer drugs.^{84–87} Some hospitals have implemented ongoing antibiotic management programs in which patients receiving certain antibiotics have their therapy reviewed by an infectious disease physician, who makes treatment recommendations to the prescribing physician when therapy is deemed to be suboptimal.⁸⁸

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY RESEARCH

While the entire process of DUR can be seen as an application of pharmacoepidemiology, the primary role for pharmacoepidemiology *research* in the area of DUR is to evaluate the effectiveness of these programs in meeting their stated goals. Namely, to what degree do they change prescribing, improve clinical outcomes, and reduce costs?

Clarifying the mechanisms by which DUR acts is also of substantial interest. For example, if DUR is effective in altering prescribing, does it do so by changing prescribing for patients identified in alerts, or by changing prescribing in future patients ("spillover" effects)? What features of DUR programs influence their effectiveness, and how might programs be made more effective?

A number of papers^{12, 89–97} have described comprehensive DUR programs or individual drug use audits without evaluating their effects. Several papers^{98–105} have examined physician and pharmacist responses to DUR alerts and attitudes regarding DUR programs, but provide no evidence regarding the effectiveness of the programs. Three uncontrolled studies,^{106–108} one study with an inappropriate control group,¹⁰⁹ and one study that failed to use the intention-to-treat principle¹¹⁰ have examined the effects of DUR programs, but are all uninterpretable because of methodologic flaws.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY RESEARCH

BARRIERS TO RANDOMIZED TRIALS

There are a number of challenges to studying the effectiveness of DUR programs. Although it is well recognized that randomized experiments provide the most convincing evidence for the effectiveness of an intervention, significant barriers exist to the

conduct of randomized trials to evaluate DUR programs. One important barrier is that in the US, Medicaid programs are required by federal law to perform DUR, making randomized trials of DUR programs impossible within Medicaid.

A barrier to randomized trials outside of Medicaid is the fact that some agencies appear interested in DUR primarily as a means to reduce medication costs, and thus have little or no incentive to study the health effects of DUR. Another major barrier is that some private sector programs consider both DUR criteria and the data needed to measure effectiveness to be proprietary information, and thus secret.¹¹¹ Despite these barriers, one experimental trial of an overall DUR program, and a number of experimental trials of individual drug use audits, have been conducted, although most have looked at drug utilization rather than health end points.

CLUSTERED OBSERVATIONS

When randomized experiments have been performed, the need to keep the intervention and control groups free from cross-contamination has necessitated the randomization of *clusters*, or units such as physicians, pharmacies, or geographic regions, rather than the randomization of individual patients. Randomization by cluster results in fewer randomized units, and thus reduced statistical power.^{112–121} However, unless clustering is correctly accounted for in the analysis, the results will tend to overstate the true precision of the data. This means that ignored clustering can artificially reduce *p*-values and the width of confidence intervals. Although statistical methods to account for clustering have been developed,^{115, 116} they have not been used widely in studies of DUR. An additional disadvantage of randomization by cluster is that clustered data techniques currently exist for relative effect measures (such as the odds ratio and the risk ratio), but not for absolute measures (such as the risk difference).

RARE OUTCOMES

An additional challenge to research in this area is that the serious health outcomes that DUR

programs are intended to prevent (for example, serious gastrointestinal bleeding caused by non-steroidal anti-inflammatory drugs) are relatively infrequent on a population basis. Thus, in most settings, the relatively small effect that DUR programs are likely to have on clinical outcome measures such as all-cause and even cause-specific hospitalizations is likely to be overwhelmed by underlying variability in the rates of these events. Therefore, studies attempting to show clinical effects of DUR programs need to be enormous in size.

POTENTIALLY NONCOMPARABLE STUDY GROUPS

Challenges associated with the use of nonrandomized study designs to make causal inferences are a central theme throughout pharmacoepidemiology (see Chapters 2, 34, 43 and 44). Not surprisingly, they are central to the interpretation of nonrandomized studies of DUR programs, as well. That is, the validity of any nonrandomized study is subject to threat by many factors, the most prominent of which is unmeasured differences between treated and untreated subjects.

Nonrandomized studies of programmatic interventions like DUR can be described using epidemiologic vocabulary (e.g., ecologic study, cohort study) as described in Chapter 2, or that of the program evaluation (“quasi-experimentation”) literature (e.g., interrupted time series, pre-post with comparison group) as developed by Campbell and Stanley¹²² and Cook and Campbell,¹²³ and summarized in Chapter 30. Because program evaluation terminology is more descriptive in the current context, it will be used here to describe nonexperimental studies of DUR programs. As described elsewhere^{122, 123} and in Chapter 30, there is a general hierarchy that ranks quasi-experimental designs with respect to scientific rigor and the strength of the causal inferences that can be drawn from them.

AVAILABILITY OF DATA

An additional methodologic challenge has arisen from the growth in managed care within US

Medicaid programs. Because US Medicaid managed care programs frequently do not report individual-level claims data back to the state Medicaid agency, these data become unavailable to investigators conducting research using data obtained from the state agency. A related challenge is the administrative and logistic barriers faced by independent researchers in gaining access to data held by private and government agencies. Decisions regarding data access need to be made with proper consideration to issues of public health, patient confidentiality, proprietary concerns, and costs of transferring data (also see Chapter 26).

CURRENTLY AVAILABLE SOLUTIONS

FRAMEWORK FOR EVALUATING THE EFFECTIVENESS OF DUR

As noted above, the true effectiveness of the DUR model can be evaluated only by examining the effects of real-life DUR programs. Thus, it is possible that DUR is ineffective *as currently implemented*, yet that it might be effective if it were implemented differently, or under different circumstances.

Real-world DUR programs operate by performing individual drug use audits, which are either discrete or ongoing. In this context, each criterion or small group of related criteria used by a DUR program is 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) will be more specific, and thus more credible than evaluations of overall DUR programs.

Similarly, evidence for more specific, proximal effects of a DUR interventions, such as changes in prescribing, tends to be more credible than evidence for more distal effects such as reductions in the rate of all-cause hospital admission. This is because more distal outcomes have greater potential to be affected by external factors.

In addition to being more credible, demonstrating an effect of DUR on prescribing is also *easier*

than showing an effect on clinical outcome measures. This is because prescribing outcomes occur much more frequently than the clinical outcomes that DUR programs are intended to prevent.

However, this increased credibility and relative ease in showing an effect comes at the price of reduced importance. That is, an effect on prescribing, even if it is the desired effect, may not result in an improvement in clinical outcomes. Further, many DUR interventions are intended not to improve clinical outcomes, but to reduce drug expenditures. Given these considerations, decision makers evaluating a given DUR program would probably be best served by research that evaluates a balance of important, albeit difficult to demonstrate clinical effects, and mechanistic evidence in the form of effects on prescribing.

Because one of the goals of DUR programs is to reduce health care costs, their ability to do so is frequently evaluated.¹²⁴ In fact, every US Medicaid program is required to estimate the cost savings of their retrospective and prospective DUR programs each year.²¹ Naturally, however, if DUR programs are ineffective in changing prescribing, they could not possibly be cost saving. If they are effective in changing prescribing, then they might result in lower drug costs. Of course, any drug cost savings could easily be overwhelmed by even small beneficial or deleterious effects on health outcomes. Thus, until the health effects of DUR programs are known, their true economic effects cannot be known.

Inpatient DUR programs often use approaches to change prescribing that are not typically used by outpatient programs, including special order forms, formulary restrictions, restriction of drugs to particular clinical specialties, and influence by local opinion leaders.²³ The effectiveness of these interventions in changing prescribing is reviewed in detail in Chapter 30. In general, these approaches can be successful in changing prescribing, although studies evaluating their clinical effects are lacking.

Nearly all outpatient DUR programs attempt to change medication use by alerting the physician of record (either directly or by alerting the dispensing pharmacist) to the existence of patient-specific exceptions to pre-established criteria. Because this

strategy is so overwhelmingly used by outpatient DUR programs, studies evaluating the effectiveness of this approach are reviewed here in detail. In particular, we will review the evidence evaluating the effect of such interventions in changing prescribing and in improving clinical outcomes.

EVIDENCE FOR THE EFFECTIVENESS OF OUTPATIENT RETROSPECTIVE DUR PROGRAMS IN CHANGING PRESCRIBING FOR THE PATIENT IDENTIFIED IN THE ALERT

Zimmerman *et al.*¹²⁵ used a nonrandomized, pre-post design with control group to evaluate the effectiveness of an intervention to reduce overuse of histamine-2 receptor antagonists (H2RAs) within the Wisconsin Medicaid DUR program. Patients who received acute dose therapy for more than 90 days were assigned to intervention or control status based on whether their prescribing physician cared for three or more such patients (intervention group) or two or fewer such patients (control group). The intervention consisted of a packet mailed to the prescriber that included a letter explaining the concerns of chronic full-dose H2RA therapy, profiles of patients with an exception, and literature references. H2RA use was measured six months pre- and 12 months postintervention. Although treatment assignment was based on physician (i.e., cluster of patients), clustering was ignored in the analysis. Therefore, the precision of the effect measures presented by the authors may overestimate the true precision.

Both study groups reduced their use of H2RAs during the 12 month postintervention period. Among patients in long term care facilities, the pre-post reduction was 16% greater in the intervention group than in the control group, although this difference was not statistically significant. Among ambulatory patients, the pre-post reduction was 30% greater in the intervention group than in the control group, a difference that was statistically significant. Thus, despite possibly incomparable treatment and control groups, and potential overstatement of the degree of precision, this study provides some evidence that a retrospective drug use audit aimed at reducing chronic

use of acute-dose H2RA therapy can have a small to moderate-sized effect on prescribing for identified patients.

Okano and Rascati¹²⁶ performed a randomized trial to examine the effectiveness of an intervention aimed at reducing duplicative peptic ulcer therapy within the Texas Medicaid population. The intervention consisted of a mailed alert letter that included a patient profile. Physicians practicing within a given small town were randomized as a single unit. Thus, randomization was by cluster, although the analysis ignored clustering, which may have resulted in overstatement of the precision of the results. The outcome measure was the continued presence of duplicative therapy six months after the intervention.

The desired outcome (discontinuation of at least one component of the duplicative therapy) was seen in 42% of patients in the intervention group and 33% of patients in the control group, for a risk ratio of the desired outcome of 1.70 (95% confidence interval, 1.12 to 2.59). Thus, these data appear to support the conclusion that a drug use audit that uses a mailed alert letter can have a modest effect on duplicative therapy. However, the true precision of the effect estimate may be less than that reported by the investigators.

Farris *et al.*¹²⁷ studied the effectiveness of a DUR intervention designed to reduce the use of broad spectrum antibiotics and first-line use of non-sedating antihistamines in a Health Maintenance Organization (HMO). The intervention group consisted of physicians at two family practice clinics, and two control groups were assembled from the remaining pool of physicians within the HMO. Thus, there were only three experimental units, and there appeared to be substantial differences in baseline characteristics among the three. Because of these limitations, the study results are difficult to interpret.

Collins *et al.*¹²⁸ performed an experimental study to evaluate the effectiveness of an intervention designed to reduce dipyridamole use among patients not receiving concomitant warfarin. The rationale was that the only FDA approved indication for oral dipyridamole was adjunctive use with warfarin in prophylaxis of thromboembolism after cardiac valve replacement. The intervention

consisted of a mailed alert letter containing patient profiles, and was performed within the Wisconsin Medicaid DUR program. For purposes of the study, the state was divided into four geographic regions, with each region assigned to one of the following study groups: no intervention, alert letters mailed to pharmacies, alert letters mailed to physicians, and alert letters mailed to both physicians and pharmacies. For practical reasons, one region was assigned to the control group, while the remaining three regions were assigned using random allocation. Subjects were followed for six months following the intervention. The proportion of patients in whom diprydamole was discontinued was compared among study groups. Clustering by region was ignored in the analysis.

In patients residing in long term care facilities, the adjusted odds ratios for discontinuation of therapy (with 95% confidence intervals) compared with the control group were: pharmacy only, 0.92 (0.55–1.55); physician only, 1.31 (0.83–2.08); pharmacy + physician, 2.10 (1.63–3.68). In ambulatory care patients, the corresponding odds ratios were: pharmacy only, 1.67 (0.89–3.14); physician only, 2.22 (1.16–4.24); pharmacy + physician, 3.81 (2.10–6.93). Predictably, the magnitude of effect appears highest in the group in which both the physician and pharmacist received the intervention, followed by the physician only and pharmacist only groups, in that order.

The difference between the physician + pharmacist group and the physician only group was not statistically significant. In contrast, the difference between the physician + pharmacist group and the pharmacist only group was statistically significant. Together, these data support the hypothesis that letters sent to physicians had a greater effect than those sent to pharmacists.

In summary, these data provide some support for the hypothesis that alert letters to physicians were effective in reducing the chronic use of a possibly ineffective medication. However, the true precision of the effect measures (as measured by the width of the 95% confidence intervals) may be less than that reported in the paper.

Sleath *et al.*¹²⁹ conducted a randomized trial to determine whether sending intervention materials to both pharmacies and physicians was more

effective than sending them only to physicians. However, because the authors reported only pre–post comparisons within each group, and did not include any between-group comparisons of pre–post differences, the study is essentially uninterpretable.

Smith *et al.*¹³⁰ conducted a randomized trial to measure the effect of an intervention aimed at reducing the use of five benzodiazepines for which the only US Food and Drug Administration approved indication was insomnia. This drug use audit was conducted as part of the Washington State Medicaid DUR program, and targeted patients who had received one of the target medications at a dose of at least one tablet per day for at least one year. Thus, the DUR program targeted patients whose treatment was well outside of published guidelines. The intervention consisted of a mailed alert letter to the prescriber that contained guidelines on sedative–hypnotic use, the patient profile or profiles treated by that prescriber, and physician-specific measures of sedative–hypnotic prescribing. Contamination of the control group by exposure to the intervention was avoided by randomizing physicians rather than individual patients. The effect of clustering on inferential statistics was eliminated by including in the analysis only one randomly selected patient per physician. The amount of sedative–hypnotic drug dispensed in the three-month postintervention period was the outcome measure of interest.

Patients in the intervention group reduced their use of target drugs by 28%, while those in the control group reduced their use by 8%, a difference that was statistically significant. The study results also suggested that some undesirable use of nontargeted sleep agents may have resulted. In addition, despite advice to the contrary, some prescribers appeared to have immediately discontinued sedative–hypnotic medications, potentially placing patients at risk of withdrawal. As well, anecdotal responses from physicians in the intervention group suggested that, despite the investigators' attempt to use well accepted criteria to identify a group of patients for whom therapy was clearly inappropriate, even careful discontinuation of sedative–hypnotic therapy may not have been in the best interest of some medically complicated

patients. That is, despite great care in developing apparently clearcut criteria, they may not have been valid, as defined above. Thus, this study provides strong evidence that a DUR program that identifies patients on long term therapy can have a moderate effect on prescribing and dispensing for the identified patient. It also points to the possibility that even carefully chosen criteria can have appropriate exceptions, and that unintended responses to alerts are possible.

Raisch and Sleath¹³¹ performed an experimental trial to measure the effectiveness of an intervention designed to reduce inappropriate prescription of anti-ulcer drugs, including histamine-2 receptor antagonists, in the New Mexico Medicaid DUR program. The intervention consisted of a mailed alert letter containing a fact sheet on prescription of anti-ulcer agents, and patient profiles. Contamination was avoided by excluding patients who were seen by physicians in both study groups. Dispensing of anti-ulcer agents in the three-month postintervention period was measured.

Forty-three percent of patients in the intervention group had no exceptions to anti-ulcer criteria in the followup period, compared to 28% in the control group (unadjusted odds ratio for the desired outcome = 1.98; 95% confidence interval, 1.23–3.18). Anti-ulcer agents were not dispensed during the followup period for 33% of intervention patients and 18% of control patients (odds ratio = 2.29; 95% confidence interval, 1.35–3.87). Clustering of patients within physician practice was ignored in the analysis, which may have artificially increased apparent precision. Regardless, this study provides relatively strong evidence that a DUR program using mailed alert letters can have a moderate effect on prescribing of chronically administered medications for patients identified in the alert letters.

A report by Abt Associates Inc.²⁸ presented results from four studies that evaluated the hypothesis that alert letters sent to prescribers can affect the drug therapy of patients identified in alert letters. None of these studies has appeared in the peer-reviewed literature as of the date of this writing. According to the report, the results suggested that patient-specific DUR intervention letters sent to prescribers were effective in (i)

increasing the prescription of misoprostil to patients receiving nonsteroidal anti-inflammatory agents, (ii) increasing the use of short acting bronchodilators in patients receiving salmeterol (a long acting β -agonist bronchodilator), and (iii) reducing long term therapy with acute-dose ranitidine. However, the level of clarity of the Abt report is not as high as the standard set by high quality peer reviewed journals. Therefore, critical evaluation of these studies is difficult.

In summary, the available evidence indicates that alert letters containing patient-specific profiles mailed to physicians can have a small but measurable effect on prescribing practices for the patient identified in the alert. There is some evidence suggesting that sending a similar letter to the pharmacy of record augments the effect of the letter to the prescriber, but this evidence is equivocal. Many published outpatient DUR efforts have focused on cost rather than quality issues. Failure to account for clustering in the analysis of studies with clustered treatment assignment is common. Publication of original research findings in peer reviewed journals should be encouraged by those who fund research.

EVIDENCE FOR THE EFFECTIVENESS OF OUTPATIENT RETROSPECTIVE DUR PROGRAMS TO CHANGE DRUG THERAPY IN PATIENTS NOT IDENTIFIED IN ALERTS ("SPILLOVER" EFFECTS)

No studies yet published in the peer reviewed literature have evaluated the impact of DUR intervention letters on prescribing for patients who are not specifically identified in the alert letters, or "spillover" effects. Three studies described in the report by Abt Associates²⁸ evaluated the hypothesis that DUR intervention letters can affect prescribing for patients other than those who were identified in alert letters. The report concluded that there was no evidence of spillover effects of an intervention designed to increase the prescription of misoprostil in patients receiving nonsteroidal anti-inflammatory agents, and in an intervention designed to increase prescription of short acting bronchodilators in patients receiving salmeterol. In contrast, the report concluded that

there was evidence of a spillover effect of an intervention designed to reduce long term therapy with acute-dose ranitidine. However, given the absence of studies appearing in the peer reviewed literature, the difficulty in evaluating the studies presented in the Abt report, and the conflicting findings within that report, it is difficult to draw any conclusions regarding the potential spillover effects of DUR programs. This remains an important, albeit challenging area for research.

EVIDENCE FOR THE EFFECTIVENESS OF OUTPATIENT RETROSPECTIVE DUR PROGRAMS IN IMPROVING CLINICAL OUTCOMES

Jay *et al.*¹³² conducted a controlled trial of the effectiveness of the California Medicaid DUR program. The study compared two counties that implemented retrospective DUR to two counties that did not, over a 24 month pre-intervention period and a 12 month post-intervention period. It examined trends in indices of health service utilization such as hospital admissions, physician visits, and drug expenditures in the Medicaid population. The study did not assess the program's effectiveness in altering prescribing. The investigators observed no materially important or statistically significant changes in health services utilization attributable to the program. Although these results suggest that the California program may have been ineffective, the magnitude of effect that can be statistically excluded is not provided, nor is it readily calculable from the data presented. The results of this study have also not been reported in the peer reviewed literature. Given that the study was performed within a single state, and examined the program's effect in the overall population rather than in any of the specific subgroups of the population that were most likely to benefit (such as those identified in an exception or those at highest risk of hospitalization), it is likely that the results can exclude the existence of only large effects. In addition, the investigators cite a number of weaknesses of the DUR program (such as the unavailability of diagnosis data, the small proportion of exceptions that resulted in alerts, and the inability to identify prescribers in

about one-fourth of exceptions) that may also have contributed to the absence of demonstrable effect.

The previously described study by Zimmerman *et al.*¹²⁵ of a DUR intervention designed to reduce overuse of histamine-2 receptor antagonists (H2RAs) also looked for an increase in the occurrence of hospitalization for gastrointestinal ulcer or bleeding, which might occur if the intervention resulted in the discontinuation of H2RAs in patients who actually needed them. The investigators did not detect an effect of the DUR intervention on this outcome, although the study was large enough only to detect a dramatically large effect. Thus, this study provides some evidence that a DUR intervention that aimed to reduce H2RA use did not enormously increase the occurrence of serious gastrointestinal ulcer and bleeding.

Thus, three decades after outpatient DUR was first proposed, and despite the existence of a cadre of thoughtful and well intentioned professionals conducting DUR programs, there is no empiric evidence supporting the effectiveness of outpatient retrospective DUR in achieving its primary objective of improving clinical outcomes. This deficiency has received substantial attention in the literature,^{3,5} and is currently being addressed by a controlled nonexperimental study funded by the National Institute on Aging (R01-AG14601), the Agency for Health Policy and Research (F32 HS00105), and the Pharmark Corporation.

EVIDENCE FOR THE EFFECTIVENESS OF OUTPATIENT PROSPECTIVE DUR PROGRAMS IN CHANGING DRUG THERAPY AND IMPROVING CLINICAL OUTCOMES

Chrischilles *et al.*^{28,133} performed a randomized trial to evaluate the effectiveness of an on-line prospective drug utilization review program. The results of this trial have not yet appeared in the peer reviewed literature. The study was conducted within the Iowa Medicaid program, and recruited pharmacies through newsletters, recruitment letters, and articles in the state pharmacy association journal. Eligible pharmacies were grouped into clusters based on shared prescribers in order to

reduce contamination of study groups. Clusters were randomized to intervention or control status. Pharmacies in the intervention group received on-line, point-of-service DUR messages that resulted from computerized reviews of prescription medication (but not diagnosis) data. Pharmacies in the control group received no on-line messages, but still could have received alerts from in-pharmacy computer systems. Pharmacies in both groups documented nondispensing activities ("cognitive services") performed for Medicaid beneficiaries.

The rate of cognitive services was not statistically different between groups (Chrischilles E, personal communication, 2 April 1999). The investigators also evaluated the effect of the intervention on (i) use of health care services, (ii) use and cost of prescription drugs, (iii) occurrence of subsequent exceptions, and (iv) occurrence of adverse events related to use of a particular set of prescription drugs, but found no consistent evidence that the intervention had an effect on any of these measures.

Thus, the conclusions do not appear to support the hypothesis that providing pharmacists with computerized DUR messages based on patients' overall prescription profile impacts drug therapy or clinical outcomes. Importantly, this study did not compare prospective DUR with no prospective DUR, but rather compared centralized prospective DUR with in-pharmacy prospective DUR (which is the current standard of care in pharmacy practice).⁴ Thus, the incremental benefit of incorporating data from other pharmacies was evaluated. Finally, it remains possible that a system that uses diagnosis data derived from medical claims to identify drug-disease interactions and omitted-but-necessary therapy might have a demonstrable effect. This hypothesis deserves to be evaluated empirically.

Monane *et al.*¹³⁴ evaluated a prospective DUR program performed by 13 mail order pharmacies owned by a pharmaceutical benefit management company. The DUR program identified prescriptions that were presented for elderly patients that violated one of 11 prespecified clinical criteria. The intervention consisted of attempted telephone contact of the prescriber, with a pharmacist initiated discussion of therapeutic alternatives.

Pharmacists performing the telephone intervention successfully contacted the prescriber in 56% of cases (95% confidence interval, 56–57%), and were successful in persuading the prescriber to change the target prescription in 8% of all cases (95% confidence interval, 8–9%). However, the "DUR change rate" of 24% presented in the paper is overly optimistic, since it (i) includes as successes cases in which the prescriber did not change the prescription, but rather stated an intention to review the patient's therapy at the next visit, and (ii) included in the denominator only cases in which the physician was successfully contacted, rather than all attempted contacts. Because no concurrent control group was included, it is not known how many prescriptions would have been intervened upon by pharmacists in the normal course of pharmacy practice, and how many of the physicians would have stopped the therapy on their own.¹³⁵ Therefore, the degree to which this program represents an incremental benefit over standard pharmacy practice is unknown.

SUMMARIZING THE EVIDENCE

Retrospective and prospective DUR are both mandated in the US for Medicaid programs, and are widely employed outside of Medicaid. Although there is evidence that some forms of retrospective DUR can have a measurable but modest impact on prescribing, there is no evidence that it achieves its primary objective of improving clinical outcomes. Further, some evidence suggests that even thoughtfully considered and well intentioned interventions aimed at improving care might have unintended consequences. For example, some physicians in the study by Smith *et al.* appeared to have abruptly discontinued sedative-hypnotic medications, which can precipitate withdrawal reactions.¹³⁰ Similarly, if there are patients who benefit from long term therapy with histamine H2 receptor antagonists, then interventions to reduce this treatment may cause instances of gastrointestinal ulcer or bleeding. Despite the intuitive appeal of prospective DUR, there is currently no empiric evidence that it provides an incremental benefit over standard pharmacy practice. Although the conduct of rigorous re-

search in this area is challenging, continued support for these programs as a means to improve clinical care should be based on empiric evidence. Three decades after outpatient DUR programs were first proposed, such evidence is still lacking.

THE FUTURE

The 1969 US DHEW Task Force on Prescription Drugs believed that DUR held great promise for improving patient care, but that evidence concerning its effectiveness was needed.^{10,11} Because DUR was adopted widely in the absence of empiric evidence, there is still little known about the effectiveness of DUR programs in improving health outcomes. Decisions to continue DUR programs should be conditional on evidence regarding its effectiveness, some of which is now being produced.

The development that is likely to have the most substantial and permanent impact on DUR in the immediate future is direct computer entry of prescriptions by the prescriber (also see Chapter 30).^{31,43} DUR systems applied in this setting will provide direct, immediate feedback to the person who is in the best position to alter therapy. Such systems would optimally incorporate drug, medical, laboratory, and formulary information in generating real-time alerts.³¹ Based on current experience, such a system would need to have a sufficiently low false positive rate so as not to foster disregard of alerts. Development of such systems is a formidable challenge, and will require ongoing, perhaps perpetual development.

Future developments will undoubtedly show even greater promise than existing approaches. However, like our drugs, these programs should be held to high standards of evidence for demonstrating that they achieve their intended goal (i.e., they are effective) and that they are free of important and predictable unintended effects (i.e., they are safe) before undergoing widespread adoption.

ACKNOWLEDGEMENT

This work was supported by grant R01-AG14601 from the National Institute on Aging.

REFERENCES

1. Manasse HR Jr. Medication use in an imperfect world: drug misadventuring as an issue of public policy, part 1. *Am J Hosp Pharm* 1989; **46**: 929–44.
2. Manasse HR Jr. Medication use in an imperfect world: drug misadventuring as an issue of public policy, part 2. *Am J Hosp Pharm* 1989; **46**: 1141–52.
3. Soumerai SB, Lipton HL. Computer-based drug-utilization review—risk, benefit, or boondoggle? *New Engl J Med*. 1995; **332**: 1641–5.
4. Lesshaft CT. Jr Ambulatory patient care. In: Osol A, Chase G, Gennaro AR, Gibson MR, Granberg CB, Harvey SC, et al., eds, *Remington's Pharmaceutical Sciences*. Easton, PA: Mack, 1980; 1631–40.
5. Lipton HL, Bird JA. Drug utilization review in ambulatory settings: state of the science and directions for outcomes research. *Med Care* 1993; **31**: 1069–82.
6. Anonymous. *Primer on Drug Use Review*. Pharmaceutical Manufacturers Association–American Medical Association–American Pharmaceutical Association, 1992.
7. Kralewski J, Wertheimer A, Ratner E. Prescription drug utilization review in the private sector. *Health Care Management Rev* 1994; **19**: 62–71.
8. Jones JK. A view from the drug utilization review management organization sector. *Clin Pharmacol Ther* 1991; **50**: 620–25.
9. Walser BL. *Pharmaceutical Cost Containment and Quality Assurance: Trends in Outpatient Utilization Review Programs and Pharmaceutical Benefits Management*. Basel: Roche, 1994.
10. Task Force on Prescription Drug Use. *Approaches to Drug Insurance Design*, Background Papers. Washington, DC: US Government Printing Office, 1969.
11. *Final Report of the Task Force on Nonprescription Drugs*. Washington, DC: US Government Printing Office, 1969.
12. Baylis RD. Drug utilization review: a description of use for a Medicaid population (Maryland) 1986–1994. *J Law Med Ethics* 1994; **22**: 247–51.
13. Knoblen JE. Drug use review. In: McLeod DC, Miller WA, eds, *The Practice of Pharmacy: Institutional and Ambulatory Pharmaceutical Services*, 1st edn. Cincinnati, OH: Harvey Whitney, 1981; 83–93.
14. Silverman M, Lydecker M. *Drug Coverage Under National Health Insurance: Policy Options*. DHEW Publication Washington, DC: (HRA) 77-3189. US Department of Health, Education, and Welfare, 1977.
15. Brodie DC. Drug utilization review 1. Planning. *Hospitals* 1972; **46**: 103–12.

16. Ebel JA. Drug utilization review. Selective surveillance. *Hospitals* 1972; **46**: 108–14.
17. Pierpaoli PG, Bowman GK. Drug utilization review 2. Implementation. *Hospitals* 1972; **46**: 95–104.
18. Rucker TD. The role of computers in drug utilization review. *Am J Hosp Pharm.* 1972; **29**: 128–34.
19. Medicare Catastrophic Coverage Act of 1988. Public Law 100–360, 1988.
20. Medicare Catastrophic Coverage Repeal Act of 1989. Public Law 101–234, 1989.
21. Omnibus Budget Reconciliation Act of 1990. Public Law 101–508, 1990.
22. Erwin WG. The definition of drug utilization review: statement of issues. *Clin Pharmacol Ther* 1991; **50**: 596–9.
23. Hennessy S, Strom BL. Issues in hospital-based drug use evaluation. *P&T: Peer-Rev J Formulary Management* 1992; **17**: 733–54.
24. Moore WJ. Medicaid drug utilization review: a critical appraisal. *Med Care Rev* 1994; **51**: 1–37.
25. Beta-Blocker Heart Attack Research Group. A randomized trial of propranolol in patients with acute myocardial infarction I. Mortality results. *J Am Med Assoc* 1982; **247**: 1707–14.
26. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *New Engl J Med.* 1992; **327**: 248–54.
27. The Norwegian Multicenter Study. Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction: a randomized trial of propranolol in patients with acute myocardial infarction I. Mortality results. *New Engl J Med* 1981; **304**: 801–7.
28. Abt Associates. *Evaluation of Drug Use Review Demonstration Projects: Final Report.* Cambridge, MA: Abt, 1998.
29. Avorn J, Soumerai SB. Use of a computer-based Medicaid drug data to analyze and correct inappropriate medication use. *J Med Syst* 1982; **6**: 377–86.
30. Avorn J, Soumerai SB. Improving drug-therapy decisions through educational outreach. A randomized controlled trial of academically based “detailing.” *New Engl J Med.* 1983; **308**: 1457–63.
31. Schiff GD, Rucker TD. Computerized prescribing: building the electronic infrastructure for better medication usage. *J Am Med Assoc* 1998; **279**: 1024–9.
32. Cataldo R, Jr. OBRA '90 and your pharmacy computer system. *Am Pharm* 1992; **NS32**: 39–41.
33. Braunstein ML, James JD. The emerging role of the computer in community pharmacy. *J Fam Pract* 1978; **7**: 1231–2.
34. Edmonds JL. Community pharmacy and computer services. *Am J Pharm Sci Support Public Health* 1972; **144**: 82–7.
35. Mars PH, van der Werf N. The computer in pharmacy II. *Pharm Weekbl* 1969; **104**: 1333–6 (in Dutch).
36. Chatellier G, Colombet I, Degoulet P. An overview of the effect of computer-assisted management of anticoagulant therapy on the quality of anticoagulation. *Int J Med Inf* 1998; **49**: 311–20.
37. Fitzmaurice DA, Hobbs FD, Delaney BC, Wilson S, McManus R. Review of computerized decision support systems for oral anticoagulation management. *Br J Haematol* 1998; **102**: 907–9.
38. Donald JB. Prescribing costs when computers are used to issue all prescriptions. *Br Med J* 1989; **299**: 28–30.
39. Rotman BL, Sullivan AN, McDonald TW, Brown BW, DeSmedt P, Goodnature D, et al. A randomized controlled trial of a computer-based physician workstation in an outpatient setting: implementation barriers to outcome evaluation. *J Am Med Inf Assoc* 1996; **3**: 340–8.
40. Vedsted P, Nielsen JN, Olesen F. Does a computerized price comparison module reduce prescribing costs in general practice? *Fam Pract* 1997; **14**: 199–203.
41. Vedsted P, Nielsen JN, Olesen F. Does a computerized price comparison module reduce prescribing costs in general practice? *Ugeskr Laeger* 1998; **160**: 4873–5 (in Danish).
42. Walton RT, Gierl C, Yudkin P, Mistry H, Vessey MP, Fox J. Evaluation of computer support for prescribing (CAPSULE) using simulated cases. *Br Med J* 1997; **315**: 791–5.
43. 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. *J Am Med Assoc* 1998; **280**: 1311–6.
44. Soumerai SB, Avorn J. Efficacy and cost-containment in hospital pharmacotherapy: state of the art and future directions. *Milbank Mem Fund Q Health Soc* 1984; **62**: 447–74.
45. Aseffa A, Desta Z, Tadesse I. Prescribing pattern of antibacterial drugs in a teaching hospital in Gondar, Ethiopia. *East Afr Med J* 1995; **72**: 56–9.
46. Ashkenazi S, Samra Z, Konisberger H, Drucker MM, Leibovici L. Factors associated with increased risk in inappropriate empiric antibiotic treatment of childhood bacteraemia. *Eur J Pediatr* 1996; **155**: 545–50.
47. Beidas S, Khamesian M. Vancomycin use in a university medical center: comparison with hospital infection control practices advisory committee guidelines [letter]. *Infect Control Hosp Epidemiol* 1996; **17**: 773–4.

48. Belliveau PP, Rothman AL, Maday CE. Limiting vancomycin use to combat vancomycin-resistant *Enterococcus faecium*. *Am J Health Syst Pharm* 1996; **53**: 1570–5.
49. Cira LM, Briceland LL, Lesar TS. Pitfalls of using antimicrobial order forms to evaluate drug use [letter]. *Am J Hosp Pharm* 1970; **51**: 700.
50. DesHarnais S, Simpson KN, Paul JE. Variations in practice patterns: antiviral drug use in hospitalized patients with herpes infections. *Am J Med Qual* 1996; **11**: 33–42.
51. Evans ME, Kortas KJ. Vancomycin use in a university medical center: comparison with hospital infection control practices advisory committee guidelines. *Infect Control Hosp Epidemiol* 1996; **17**: 356–9.
52. Fonseca SN, Ehrenkranz RA, Baltimore RS. Epidemiology of antibiotic use in a neonatal intensive care unit. *Infect Control Hosp Epidemiol* 1994; **15**: 156–62.
53. Fraser GL, Stogsdill P, Dickens JDJ, Wennberg DE, Smith RPJ, *et al*. Antibiotic optimization. An evaluation of patient safety and economic outcomes. *Arch Intern Med* 1997; **157**: 1689–94.
54. Goeckner BJ, Hendershot E, Scott K, Drake M. A vancomycin monitoring program at a community hospital. *Joint Committee J Qual Improv* 1998; **24**: 379–85.
55. Gutierrez F, Wall P, Cohen J. An analysis of the trends in the use of antifungal drugs and fungal isolates in a UK University Hospital [letter]. *J Hosp Infect* 1995; **31**: 149–52.
56. Harbarth S, Pittet D, Gabriel V, Garbino J, Lew D. Cefepime—assessment of its need at a tertiary care center. *J Clin Pharm Ther* 1998; **23**: 11–7.
57. Janknegt R, Wijnands WJ, Caprasse M, Brandenburg W, Schuitenmaker MG, Stobberingh E. Antimicrobial drug use in hospitals in The Netherlands, Germany and Belgium. *Eur J Clin Microbiol Infect Dis* 1993; **12**: 832–8.
58. Jozefiak ET, Lewicki JE, Kozinn WP. Computer-assisted antimicrobial surveillance in a community teaching hospital. *Am J Health Syst Pharm* 1995; **52**: 1536–40.
59. Lee KR, Leggiadro RJ, Burch KJ. Drug use evaluation of antibiotics in a pediatric teaching hospital. *Infect Control Hosp Epidemiol* 1994; **15**: 710–2.
60. LeMire M, Wing L, Gordon DL. An audit of third generation cephalosporin prescribing in a tertiary care hospital. *Aust N Z J Med* 1996; **26**: 386–90.
61. Logsdon BA, Lee KR, Luedtke G, Barrett FF. Evaluation of vancomycin use in a pediatric teaching hospital based on CDC criteria. *Infect Control Hosp Epidemiol* 1997; **18**: 780–2.
62. Natsch S, Conrad C, Hartmeier C, Schmid B. Use of amoxicillin-clavulanate and resistance in *Escherichia coli* over a 4-year period. *Infect Control Hosp Epidemiol* 1998; **19**: 653–6.
63. Rahal JJ, Urban C, Horn D, Freeman K, Segal-Maurer S, Maurer J, *et al*. Class restriction of cephalosporin use to control total cephalosporin resistance in nosocomial *Klebsiella*. *J Am Med Assoc* 1998; **280**: 1233–7.
64. Redding PJ, Taylor A, Smith AM. Audit of parenteral antimicrobial agents and infusion fluids. *Health Bull* 1995; **53**: 110–4.
65. Rovers JP, Bjornson DC. Assessment of methods and outcomes: using modified inpatient ciprofloxacin criteria in community-based drug use evaluation. *Ann Pharmacother* 1994; **28**: 714–9.
66. Singer MV, Haft R, Barlam T, Aronson M, Shafer A, Sands KE. Vancomycin control measures at a tertiary-care hospital: impact of interventions on volume and patterns of use. *Infect Control Hosp Epidemiol* 1998; **19**: 248–53.
67. Soto J. Strategies for rationing the use of antibiotics in hospitals. *Enferm Infecc Microbiol Clin* 1996; **14**: 576–7 (in Spanish).
68. Speirs GE, Fenelon LE, Reeves DS, Speller DC, Smyth EG, Wilcox MH, *et al*. An audit of ciprofloxacin use in a district general hospital. *J Antimicrob Chemother* 1995; **36**: 201–7.
69. Thomas M, Govil S, Moses BV, Joseph A. Monitoring of antibiotic use in a primary and tertiary care hospital. *J Clin Epidemiol* 1996; **49**: 251–4.
70. van der Hoek W. Prescription audit for antibiotics in a district hospital [letter]. *Trop Doct* 1994; **24**: 85–6.
71. Wright SW, Wrenn KD. Appropriateness of vancomycin use in the emergency department. *Ann Emerg Med* 1998; **32**: 531–6.
72. Zadik PM, Moore AP. Antimicrobial associations of an outbreak of diarrhoea due to *Clostridium difficile*. *J Hosp Infect* 1998; **39**: 189–93.
73. Zaran FK, Weingarten CM, Rybak MJ, Stevenson JG. Effect of constructive feedback on pharmacist handling of orders for monitored antimicrobials. *Am J Health Syst Pharm* 1996; **53**: 2196–8.
74. Dasta JF, Fuhrman TM, McCandles C. Patterns of prescribing and administering drugs for agitation and pain in patients in a surgical intensive care unit. *Crit Care Med* 1994; **22**: 974–80.
75. Dasta JF, Fuhrman TM, McCandles C. Use of sedatives and analgesics in a surgical intensive care unit: a follow-up and commentary. *Heart Lung* 1995; **24**: 76–8.
76. Friedland LR, Kulick RM. Emergency department analgesic use in pediatric trauma victims with fractures. *Ann Emerg Med* 1994; **23**: 203–7.
77. Tholl DA, Wager MS, Sajous CH, Myers TF. Morphine use and adverse effects in a neonatal

- intensive care unit. *Am J Hosp Pharm* 1994; **51**: 2801-3.
78. Turner R, Clark J, Root T, McElligott L, Hardy J. An audit of morphine prescribing in a specialist cancer hospital. *Palliat Med* 1994; **8**: 5-10.
 79. Woolard DJ, Terndrup TE. Sedative-analgesic agent administration in children: analysis of use and complications in the emergency department. *J Emerg Med* 1994; **12**: 453-61.
 80. Zisselman MH, Rovner BW, Kelly KG, Woods C. Benzodiazepine utilization in a university hospital. *Am J Med Qual* 1994; **9**: 138-41.
 81. Alran C, Damase-Michel C, Celotto N, Durand MC, Montastruc JL. Consumption of benzodiazepines in a French university hospital between 1980 and 1991. *Fundam Clin Pharmacol* 1993; **7**: 319-23.
 82. Stone P, Phillips C, Spruyt O, Waight C. A comparison of the use of sedatives in a hospital support team and in a hospice. *Palliat Med* 1997; **11**: 140-4.
 83. Yuen EJ, Zisselman MH, Louis DZ, Rovner BW. Sedative-hypnotic use by the elderly: effects on hospital length of stay and costs. *J Ment Health Adm* 1997; **24**: 90-7.
 84. Wyncoll DL, Roberts PC, Beale RJ, McLuckie A. H2 blockers in the intensive care unit: ignoring the evidence? Telephone survey. *Br Med J* 1997; **314**: 1013.
 85. Barnhart MR. Effect of physician education on omeprazole use at a small public hospital [letter]. *Am J Health Syst Pharm* 1996; **53**: 1334.
 86. Brandhagen DJ, Pheley AM, Onstad GR, Freeman ML, Lurie N. Omeprazole use at an urban county teaching hospital. *J Gen Intern Med* 1995; **10**: 513-5.
 87. Segal R, Russell WL, Oh T, Ben-Joseph R. Use of i.v. cimetidine, ranitidine, and famotidine in 40 hospitals. *Am J Hosp Pharm* 1993; **50**: 2077-81.
 88. John JFJ, Fishman NO. Programmatic role of the infectious diseases physician in controlling antimicrobial costs in the hospital. *Clin Infect Dis* 1997; **24**: 471-85.
 89. Blackburn JL, Downey FW, Quinn TJ. The Saskatchewan program for rational drug therapy: effects on utilization of mood-modifying drugs. *Drug Intell Clin Pharm* 1990; **24**: 878-82.
 90. DeNeef P, Ellsworth A, Schneeweiss R. A system for drug utilization review in ambulatory care. *J Fam Pract* 1991; **32**: 607-12.
 91. Harley NA. Pennsylvania takes lead in drug education venture. *Pa Med* 1993; **96**: 12-4.
 92. Kaplan B, Farris KB, Kirking DM. Assessing physician choice of nonsteroidal antiinflammatory drugs in a health maintenance organization. *Ann Pharmacother* 1993; **27**: 1393-9.
 93. Lipowski EE, Bauwens SF, Collins TM. An examination of histamine-2 receptor antagonist use by Medicaid recipients in Wisconsin long-term care facilities. *J Am Geriatr Soc* 1988; **36**: 531-6.
 94. Rascati K. Drug utilization review of concomitant use of specific serotonin reuptake inhibitors or clomipramine with antianxiety/sleep medications. *Clin Ther* 1995; **17**: 786-90.
 95. Sandusky M. Michigan drug utilization review and OBRA 90. *Mich Med* 1993; **92**: 38-9.
 96. Sena MM, Pashko S. Drug utilization review using a Medicaid claims database. *Clin Ther* 1993; **15**: 900-4.
 97. Spore DL, Mor V, Larrat P, Hawes C, Hiris J. Inappropriate drug prescriptions for elderly residents of board and care facilities. *Am J Public Health* 1997; **87**: 404-9.
 98. Armstrong EP, Proteau D. Retrospective drug utilization review software systems: perspectives of state Medicaid DUR directors. *Ann Pharmacother* 1996; **30**: 1088-91.
 99. Armstrong EP, Markson TJ. On-line DUR messages: pharmacists' attitudes and actions in response. *J Am Pharm Assoc* 1997; **NS37**: 315-20.
 100. Brown CM, Lipowski EE. Pharmacists' reactions to the Wisconsin Medicaid drug-use review program. *Am J Hosp Pharm* 1993; **50**: 1898-902.
 101. Lipowski EE, Becker M. Presentation of drug prescribing guidelines and physician response. *Qual Rev Bull* 1992; **18**: 461-70.
 102. Rascati KL, Okano GJ, Burch C. Evaluation of physician intervention letters. *Med Care* 1996; **34**: 760-6.
 103. Sandusky M. DUR intervention letters: how pharmacists respond. *Am Pharm* 1993; **NS33**: 58-64.
 104. Sleath B, McCament-Mann L, Collins T, Hollarbush J. Response forms reflect pharmacists' participation in retrospective DUR. *J Am Pharm Assoc* 1997; **NS37**: 77-84.
 105. Sleath B, Collins T. Physician responses to an educational intervention on improving their long-term prescribing of sedatives. *Patient Educ Couns* 1997; **31**: 215-22.
 106. Holm RP, Helgeland D. DEEP, the South Dakota Medicaid drug use review program. Is the hassle justified? *S D J Med* 1993; **46**: 368-9.
 107. LeGrady D. The Nebraska Medicaid Drug Utilization Review Program. *Nebr Med J* 1992; **77**: 3-6.
 108. Groves RE. Therapeutic drug-use review for the Florida Medicaid program. *Am J Hosp Pharm* 1985; **42**: 316-9.
 109. Guo JJ, Gibson JT, Hancock GR, Barker KN. Retrospective drug utilization review and the behavior of Medicaid prescribers: an empirical marginal analysis. *Clin Ther* 1995; **17**: 1174-87.
 110. Raisch DW, Hammond RW, Hollarbush JL, Wood M. Effect of a pharmacist-education initiative on ketorolac use and costs in a Medicaid

- program. *Am J Health Syst Pharm* 1998; **55**: 1274–80.
111. Lipton HL. Pharmacy benefit management companies: dimensions of performance. *Annu Rev Public Health* 1999; **20**: 361–401.
 112. Bland JM, Kerry SM. Statistics notes. Trials randomised in clusters. *Br Med J* 1997; **315**: 600.
 113. Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979–1989. *Int J Epidemiol* 1990; **19**: 795–800.
 114. Donner A. Sample size requirements for stratified cluster randomization designs. *Stat Med* 1992; **11**: 743–50.
 115. Donner A, Klar N. Confidence interval construction for effect measures arising from cluster randomization trials. *J Clin Epidemiol* 1993; **46**: 123–31.
 116. Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol* 1994; **140**: 279–89.
 117. Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. *Br Med J* 1998; **316**: 1455.
 118. Kerry SM, Bland JM. Trials which randomize practices II: sample size. *Fam Pract* 1998; **15**: 84–7.
 119. Kerry SM, Bland JM. Trials which randomize practices I: how should they be analysed?. *Fam Pract* 1998; **15**: 80–3.
 120. Kerry SM, Bland JM. Sample size in cluster randomisation. *Br Med J* 1998; **316**: 549.
 121. Kerry SM, Bland JM. Analysis of a trial randomised in clusters. *Br Med J* 1998; **316**: 54.
 122. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Boston, MA: Houghton Mifflin, 1966.
 123. Cook TD, Campbell DT. *Quasi-Experimentation: Design and Analysis for Field Studies*. Boston, MA: Houghton Mifflin, 1979.
 124. Kreling DH, Mott DA. The cost effectiveness of drug utilisation review in an outpatient setting. *Pharmacoeconomics* 1993; **4**: 414–36.
 125. Zimmerman DR, Collins TM, Lipowski EE, Sainfort F. Evaluation of a DUR intervention: a case study of histamine antagonists. *Inquiry* 1994; **31**: 89–101.
 126. Okano GJ, Rascati KL. Effects of Medicaid drug utilization review intervention letters. *Clin Ther* 1995; **17**: 525–33.
 127. Farris KB, Kirking DM, Shimp LA, Opdycke RA. Design and results of a group counter-detailing DUR educational program. *Pharm Res* 1996; **13**: 1445–52.
 128. Collins TM, Mott DA, Bigelow WE, Zimmerman DR. A controlled letter intervention to change prescribing behavior: results of a dual-targeted approach. *Health Serv Res* 1997; **32**: 471–89.
 129. Sleath B, Collins T, Kelly HW, McCament-Mann L, Lien T. Effect of including both physicians and pharmacists in an asthma drug-use review intervention. *Am J Health Syst Pharm* 1997; **54**: 2197–200.
 130. 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.
 131. Raisch DW, Sleath BL. Influencing the use of anti-ulcer agents in a Medicaid program through patient-specific prescribing feedback letters. *J Gen Intern Med* 1999; **14**: 145–50.
 132. Jay DE, Eynon BP, Javitz HS. *The Medi-Cal Therapeutic Drug Utilization Review Project*. Menlo Park, CA: SRI International, 1991.
 133. Chrischilles EA. *The Iowa Medicaid OPDUR Demonstration Project Annual Report*. Des Moines, IA: Iowa Department of Human Services, 1997.
 134. Monane M, Mathias DM, Nagle BA, Kelly MA. Improving prescribing patterns for the elderly through an online drug utilization review intervention: a system linking the physician, pharmacist, and computer. *J Am Med Assoc* 1998; **280**: 1249–52.
 135. Soumerai SB, Lipton HL. Improving prescribing patterns for the elderly through an online drug utilization review program [letter]. *J Am Med Assoc* 1999; **281**: 1168–9.

32

Determining Causation from Case Reports

JUDITH K. JONES

The Degge Group, Ltd., and Adjunct Faculty, George Washington University, Georgetown University, Washington, DC, USA

INTRODUCTION

A major component of the evaluation of reports of suspected adverse drug reactions, or events in a clinical trial, is a judgment about the degree to which any reported event is, in fact, causally associated with the suspected drug. In reality, a particular event either is associated or is not associated with a particular drug, but the current state of information almost never allows a definitive determination of this dichotomy. Accordingly, a number of approaches to the determination of the probability of a causal drug–event association have evolved over the past several years. This chapter will first discuss the historical development of these efforts, and several of the current approaches, uses, and evolving efforts will then be reviewed, including a brief consideration of the evaluation of single events in the clinical trial setting.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY RESEARCH

The basic clinical problem to be addressed is illustrated in Figure 32.1. A clinical event occurs

within the milieu of a number of possible causal factors. That event either occurred independently or in some way its occurrence was partially or totally linked to one or more of the potential causative agents. The primary task is to determine the degree to which the occurrence of the event is linked to one particular suspected causal agent, in this case a drug.

This task of evaluating causality in case reports shares some similarities with the problem of evaluating causality in chronic disease epidemiology, as discussed below and in more detail in Chapter 2. However, both the nature of the exposure and sometimes the nature of the event make the determination of causality in case reports a major challenge. This set of circumstances presents a special challenge to anyone who evaluates cases of suspected adverse drug reactions, as the evaluator must make, at the very least, an implicit judgment of causality. It also presents a special challenge to those who desire a coherent, consistent, and reliable method of determining whether there is a causal relationship between these events. Specifically:

1. The usual focus of adverse reaction assessment is an individual clinical event suspected of being

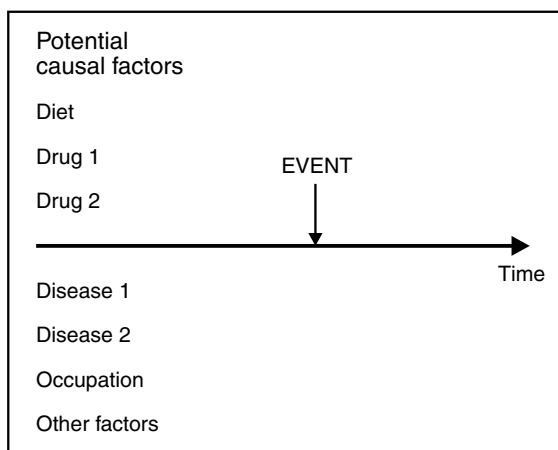


Figure 32.1. Diagram depicting the dilemma for determining causation of an event in a clinical setting. In reality a drug either did or did not cause or contribute to an event. However, given the multiple factors associated with the event, the actual truth can seldom be ascertained. Instead, some expression of probability that the drug was associated with the event is made. The method by which this expression is determined is the primary concern of those in adverse reaction causality research.

associated with a drug. The reporting of this event will, therefore, typically be in the context of a suspicion that the event is drug induced, which will often bias the collection of data required to evaluate other possible causes.

2. The data available about a patient's drug exposure in the typical case report are incomplete, usually missing precise information on duration, actual dose ingested, and/or concomitant drugs administered. The one exception is the report that comes from a hospital, but this is less often a source of reports because of the excess of events and exposures that occur during a hospitalization, confounding most suspicions.
3. The data available on the event, including its onset, characteristics, and time course, are also typically incomplete, because the suspicion is usually retrospective and the desired data (e.g., baseline laboratory data) are often not available.
4. As a corollary to 3 above, since adverse reactions can be acute, subacute, or chronic,

can be reversible or not (e.g., death and birth defects), can be rare or common, and can be unique pathologically or identical to known common diseases, it has been difficult to define general data elements and criteria which will apply to most types of adverse reaction. For example, for irreversible events, criteria that require dechallenge and rechallenge are irrelevant. More recently, as described below, this is changing.

5. Data on concomitant diseases and other confounding conditions, such as diet and habits, are typically not available, often because of factors discussed in 1 and 3 above.

Closely linked to the task of determining whether there is a causal relationship between a drug exposure and an event is the reason for that particular determination and the impact of that inference on any actions taken. If the determination is perceived to have little impact, it might logically be less rigorous. However, the evolution of recent interest in the entire subject of adverse drug reactions and the recognition of the broad use of alleged drug–event associations has supported a need for consistent and reliable methods of causality determination.¹

Additionally, with the appearance of regulations for reporting in clinical trials of adverse events that are “reasonably” associated with a drug,² there is a growing need to describe the basis for defining an association within this setting. Although some of the above-listed uncertainties are less likely to exist, there are similar difficulties with rare events, which will be considered below.

HISTORICAL PERSPECTIVES

Development of Concepts of Causality for Adverse Reactions

The development of thinking about the causality of adverse reactions has evolved in two disciplines: (i) in epidemiology, and (ii) in the study of individual case reports of adverse reactions. Consideration of both is important.

In the 1950s, epidemiologists grappled with the issue of causality, and Yerushalmy and Palmer, drawing upon the Bradford Hill epidemiology criteria as well as the Koch–Henle postulates for establishing causation for infectious diseases, developed a set of proposed criteria for causality. These evolved, after considerable deliberation with other epidemiologists, into five criteria for the causal nature of an association.³ These included determinations of:

1. the consistency of the association;
2. the strength of the association;
3. the specificity of the association;
4. the temporal relationship of the association; and
5. the coherence, or biological plausibility, of the association.

These criteria continue to be generally used in chronic disease epidemiology, although they have also been actively discussed and criticized.³ They are discussed in more detail in Chapter 2. Although seldom explicitly noted, the reasoning behind these criteria appeared at about the same time as did thinking about the causal assessment of individual reports of adverse reactions.

Prior to the last two decades in the adverse reactions field, the typical approach to case reports of suspected drug associated clinical events was to consider the events as possibly associated with the drug if there were a number of similar reports. Considerations of pharmacologic plausibility, dose–response, and timing factors were sometimes implicit, but seldom explicit. This approach continued until relatively recently, and in some cases is still used.

The more perplexing proposed drug–event associations were then typically referred to one or more experts, who generally approached the evaluation by what has been termed “global introspection.” In this approach, the experts collect all the facts relevant to the problem at hand, compile them, and make unstructured judgements to decide the answer. In the causality assessment context, this answer has usually been expressed in terms of a qualitative probability scale, for example “definite” versus “probable”

versus “possible” versus “doubtful” versus “unrelated.”⁴

The subjective nature of global reasoning as an approach led a number of investigators to develop more structured methods of causality assessment. Irely, in examining the details of cases of suspected reactions at the US Armed Forces Institute of Pathology, clearly demonstrated the discrepancy between cases initially reported as drug associated and those found by careful detailed examination to be actually likely to be drug associated.^{5,6} Shortly thereafter, the clinical pharmacologists Karch and Lasagna also recognized the inadequacy of expert “global” evaluations of adverse reactions and developed a decision table, or algorithm, to segment the evaluation of a case into several components.⁷ These two groups of investigators identified very similar basic data elements that they felt were necessary for a more standardized assessment:

1. the timing of the event, relative to the drug exposure;
2. the presence or absence of other factors which might also cause the event;
3. the result of withdrawing the drug (“dechallenge”);
4. the result of reintroducing the drug (“rechallenge”); and
5. other data supporting an association, e.g., previous cases.

Although the criteria shared some similarities with those in chronic disease epidemiology, the special characteristics of adverse drug reactions required considerations that differed somewhat. The usual setting for this latter causality assessment was a single case or group of cases from an ill defined exposed population. The criteria derived for chronic disease epidemiology were therefore inapplicable.

Following the introduction of these new methods for the assessment of suspected adverse drug reactions, a large number of other approaches were developed,^{8–15} either as algorithms, decision tables, or, in at least one case, as a diagrammatic method.¹³ These were reviewed and summarized in monographs from two con-

Table 32.1. A summary of the information categories in the method of Kramer *et al.* (8) for determining the causality of adverse reactions

Axis ^a	Information category	Number of questions ^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 issues
V	Dechallenge	23	All aspects of timing, results
VI	Rechallenge	10	All aspects of response

^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.

^bNot all questions are asked for any one problem.

ferences held in the early 1980s on the causality of adverse reactions—one in Morges, Switzerland,¹⁶ and another in Crystal City, VA.¹⁷ The vast majority of these methods shared the basic elements originally suggested by Irey and by Karch and Lasagna, but many added many other details useful for the evaluation of special cases, such as injection site reactions or *in vitro* verification (e.g., 14). Some included extensive scoring systems linked to relatively extensive algorithms, such as the approach published by Kramer *et al.*⁸ A summary of some of the major methods is presented in Table 32.1, and selected methods are discussed in detail in the next section of this chapter.

The 1981 Morges conference,¹⁶ the 1983 Crystal City conference,¹⁷ and a 1983 Paris meeting¹⁸ were all intended to compare a number of these approaches and to consider whether a single method might be developed that could represent a consensus. An international study group, the Associated Permanent Workshop of Imputologists (APWI) (“imputology” being the French term for causality), was initiated at the Morges meeting and continued into the 1990s.^{19,20} Although a consensus method was not established, the Crystal City conference had requested an outside observer (Dr. David Lane, a theoretical statistician) to provide a critique of the deliberations.²¹ His critique and subsequent participation in the Paris conference and APWI resulted in the development of a new approach for assessing the causality of adverse reactions based on Bayes probability

theorem.²² This approach considered the probability of an event occurring in the presence of a drug relative to its probability of occurring in the absence of the drug, considering all details of the case.^{23–26} Although in use elsewhere in medicine, this approach had not been applied to analyses of suspected adverse effects. This method will be discussed in more detail in the next section of this chapter.

Potential Uses of Causality Assessment

Despite the proliferation of methods and the great interest in adverse effects of drugs, the actual use of these methods for decision making has been infrequent. However, causality assessment has been required in France for many years and has been formally considered in a European Community Directive^{27,28} (see also Chapter 11). This has resulted in a general consensus on the causality terms used by the European Union member states.²⁹ In fact, there are a variety of settings where these approaches could be useful, from drug development and monitoring by the pharmaceutical manufacturer, to evaluation and monitoring by regulators, to the clinical setting, the courtroom, and even the newsroom.¹

Pharmaceutical Manufacturers

Manufacturers of pharmaceuticals must view causality assessment for events associated with their drugs from the standpoints of both regula-

tory requirements and product liability. Until recently, US pharmaceutical manufacturers have not had to consider assessment of causality for regulatory purposes. Regulations covering post-marketing event monitoring in the US required reporting of all events associated with the drug “whether or not thought to be associated with the drug” (earlier, US Code of Federal Regulations 21:310.300; now, 21:314.80) (see also Chapter 10), and causality assessment did not formally apply to events in clinical trials. However, recently US FDA regulations implicitly require causality assessment for determination of reporting certain types of clinical event in clinical trials in the Investigational New Drug (IND) regulations (CFR 21:312.22). This has come about by requiring the reporting of serious unexpected events associated with use of a drug where there is a reasonable possibility that the events may have been caused by the drug. A disclaimer is noted that such a safety report does not constitute an admission that the drug caused the events. These regulations do not provide criteria or a suggested method; however, they do imply that such methods might be useful.

Independently, manufacturers faced with a serious event in drug development have also been motivated to explore formal evaluation methods for serious events that could be used to signal the need for discontinuation of drug development if causality is established.

Outside of the US, the requirements for manufacturers to consider causality have varied from country to country. Many regulatory agencies have requested or implied some type of evaluation to minimize the number of nonspecific events reported.³⁰ Given this environment, particularly in a growing international milieu, manufacturers have been actively interested in this area. In fact, several of the specific methods for causality assessment have been published by investigators based in the pharmaceutical industry.^{11, 14, 15, 31–33}

Causality definitely is an issue for pharmaceutical manufacturers in the arena of product liability (see Chapter 9). Freilich has considered many aspects of this,³⁴ concluding that a company must have a rigorous process for the review of any adverse event reports and “make

causality assessments on an ongoing basis” for product liability purposes. This is necessary to comply with the duty to warn, which he summarizes as follows: “Information must be given of any risks of death or serious harm, no matter how rare, as well as information concerning side effects where there is a substantial probability of their occurrence, no matter how mild.” Others in the legal arena dealing in product liability have considered causality issues and the notion of the “substantial factor” test for contributing to causation.^{35, 36}

Drug Regulators

The use by drug regulators of causality assessment of spontaneously reported postmarketing adverse reactions has varied considerably. Most countries’ drug regulators have some method of approaching causality, but this method has been best defined in France, Australia, and certain other countries^{30, 37, 38} (also see Chapter 11).

In France—owing in part to the considerable original work and interest in adverse reaction causality by a regulator, J. Dangoumou, and his colleagues—all reports of suspected reactions must be evaluated by the “French method.” This method combines symptom and chronologic criteria to give a “global intrinsic score,” and then adds bibliographic data from standardized sources to give an “extrinsic score.”^{10, 39}

In the US (see also Chapter 10), although a data element for causality was incorporated into the initial file format for the computerized reports of suspected adverse reactions submitted to the FDA, no formal method for evaluating all reports was used until a simple algorithm was developed in the early 1980s, based on the Irey and the Karch and Lasagna work.^{12, 13} This method very specifically excluded the consideration of previous literature reports as a basis for considering the strength of the association. It was reasoned that, in many cases, the FDA would be in the position of receiving the first reports of an association, and such a criterion would suppress a signal of a possibly new drug associated event. The primary use of the assessment by the FDA was administrative, i.e., the causality assessment was a mechanism for identifying the best

documented cases—those with a “probable” or “highly probable” association. The causality judgment was specifically deleted from publicly available files, which consistently carry the caveat, “a cause and effect relationship has not been established.” Although the FDA algorithm existed for the reviewers of the reports, the frequency of its actual use was not determined, and this causality data element was removed from the computer file in 1986. The FDA does not now use formal causality assessment on a routine basis (see Chapters 8 and 10).

Publishers of Reports of Adverse Reactions

The medical literature containing case reports of suspected adverse reactions has largely avoided the issue of causality. The majority of single case reports, letters to the editor, or short publications do not provide an explicit judgement using any of the published algorithms. Further, many reports do not provide information on confounding drug therapy or medical conditions, data elements considered by most knowledgeable in adverse reaction assessment to be essential for considering causality. This issue was recognized as one of several problems relating to the publication of adverse reactions in the literature, and was discussed extensively during the conference in Morges, Switzerland in 1983. A number of editors of medical publications were present and discussed the quality of information in reported cases. They developed a list of the types of information that would be desirable for published reports, information that would permit the reader to assess independently the likelihood of the association.^{40,41} More recently, Harambaru and her colleagues compared the value of 500 published reports with 500 spontaneous reports with respect to the availability of information needed in most standard causality assessments. Although analysis suggested the published reports contained significantly more information, the tabulation suggests very sparse data on both alternate causes/other diseases and other drugs in *both* types of report.⁴²

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY RESEARCH

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 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 oculomucocutaneous 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 epidemiological studies or clinical trials.⁴³ 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 drug withdrawal is a consideration.⁴⁴ This latter view has spurred the evolution of causal evaluation from expert consensual opinion based on global introspection to structured algorithms, and to elaborate probabilistic approaches, as described previously. Further, because of the nature of drug associated effects, particularly those that are rare and serious, the question has been raised about whether epidemiologists need to consider using methods for causal evaluations of cases in their formal studies and in clinical trials, since the small

numbers available may not be amenable to standard statistical analysis.⁴⁵

CURRENTLY AVAILABLE SOLUTIONS

There are now a variety of methods for causality assessment of spontaneous reports. Four basic types will be described, chosen as illustrative examples and because they have been widely described in various publications.

UNSTRUCTURED CLINICAL JUDGEMENT/GLOBAL INTROSPECTION

Probably the most common approach to causality assessment is unstructured clinical judgement. An expert is asked to review the clinical information available and to make a judgement 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.⁴

First, cognitive psychologists have shown that the ability of the human brain to make unaided assessments of uncertainty in complicated situations is poor, especially when assessing the probability of a cause given an effect, precisely the task of causality assessment.⁴⁶ 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.^{11–13,47,48}

Second, global introspection is uncalibrated. One assessor’s “possible” might mean the same thing as another assessor’s “probable.” This has been well demonstrated in a study of one pharmaceutical company’s spontaneous report reviewers, who used both a verbal and numerical scale.¹⁶ These and other shortcomings of global introspection as a causality assessment method for adverse reactions are discussed in detail by Lane, Hutchinson, and Kramer.^{4,21,49}

ALGORITHM/CRITERIAL METHOD WITH VERBAL JUDGMENTS

The subsequent attempts to address this problem have resulted in the proliferation of approaches (see references 16 and 17 for reviews and examples of these methods, the appendix in reference 20 for a complete bibliography, and a recent summary).⁵⁰ These methods range from simple flow charts posing ten or fewer questions to lengthy questionnaires containing up to 84 items. However, they share a common basic structure essentially based on the original Karch and Lasagna and Irely work—the timing of the adverse event in relation to administration of the drug, alternative etiological candidates, previous recognition of the event as a possible adverse reaction to the drug, the response when the drug is discontinued (dechallenge), and the response when the drug is subsequently readministered (rechallenge). Information relevant to each factor is elicited by a series of questions, the answers to which are restricted to “yes/no” (and, for some methods, “don’t know”).

These approaches have advantages when compared to global introspection,⁴⁹ since 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 of the algorithms or decision tables. These judgments require, in some cases, “yes” or “no” answers where, in fact, a more quantitative estimate of uncertainty would be more appropriate. For example, the reviewer might have to consider whether the appearance of jaundice within one week represented a sufficient duration of drug exposure to be consistent with a drug–event association. Even adherents of some of the methods agree that their procedures for converting answers into probability ratings are arbitrary.

This type of approach, with various degrees of complexity, is used by some drug regulatory agencies, such as that of Australia.³⁷ The FDA algorithm, currently not in official use, was

another example of this approach, inquiring sequentially about temporal sequence, dechallenge, rechallenge, and concomitant diseases that might have caused the event. Based on the Irey and the Karch and Lasagna concepts, it was tailored to be amenable for rapid use by professionals with varied backgrounds for the administrative purpose of finding well documented cases for regulatory signal evaluation. It was also considered useful and easily remembered by clinicians in initial differential diagnosis of a clinical event. However, this very simple approach is less useful for irreversible drug effects, since they have neither dechallenge or rechallenge possibilities. To address this, an alternate algorithm for fatal outcome events was developed in the aftermath of the FDA algorithm (Dries, in 14).

ALGORITHMS REQUIRING SCORING OF INDIVIDUAL JUDGMENTS

Many algorithms permit quantitative judgments by requiring the scoring of their criteria. The answers to the algorithms' questions are converted into a score for each factor, the factor scores are summed, and this overall score is converted into a value on a quantitative probability scale. These judgments range from the extensive, multiple question method of Venulet *et al.*,¹⁴ which has now been translated for computer use, to the relatively simpler French

method.¹⁰ The method developed by Kramer *et al.*⁸ received considerable review and is representative of the scored methods. Although it was presented in algorithm format with multiple steps, it can also be represented in tabular format, as shown here (Table 32.1). One of the more practical methods of this type was developed by Naranjo *et al.*⁹ This has been adopted in a number of clinical settings (Naranjo, personal communication) and is shown in Figure 32.2.

These quantitative methods have found applications in a number of settings, ranging from evaluations of suspected adverse reactions by hospital committees (US hospitals are now required by the Joint Commission on Accreditation of Health Care Organizations (JCAHO) to have programs of adverse reaction surveillance) to use by some regulatory authorities, as in France. They are also used, although sometimes only in a research context, by some pharmaceutical manufacturers.^{16,30} The specific manner in which they are used has not been well described in the literature.

PROBABILISTIC METHODS

Recognition of the various problems inherent in the previously existing methods set the stage for the development of an alternative approach based on the Bayesian probability approach to

CAUSALITY ASSESSMENT EXAMPLE: 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	-
+ Rechallenge?	+2	-1	0	-
Alternative causes?	-1	+2	0	-
Reaction with placebo?	-1	+1	0	-
Drug blood level toxic?	+1	0	0	-
Reaction dose-related?	+1	0	0	-
Past history of similar event?	+1	0	0	-
ADR confirmed objectively?	+1	0	0	= -

Figure 32.2. A critical scored algorithm illustrated by the method of Naranjo *et al.*, in wide use.⁹ This particular method uses some of the basic data elements as well as more details of the history and characteristics of the case, and an arbitrary score is designated for each response.

assessment of causality. This method has provided an opportunity for a fresh look at the issue of causality, and its initial apparent difficulty (due to its requirement for using all available information) raised some new issues about causality assessment of adverse reactions. It has also brought the area of adverse reactions evaluation into a larger discussion of the value of the Bayesian and probabilistic approaches to the analysis of medical and scientific data.³

First published as a method for adverse reaction assessment by Auriche,³³ who participated with Lane and others in a working group within the APWI organization, this method was first presented in extensive form in a workshop in 1985 (see Figure 32.3). Several examples were published in a monograph and subsequently in early papers.^{25,26} The methods have been incorporated into automated versions by both Naranjo and Hutchinson, the latter developing a model using an expert system.^{51,52} Naranjo and colleagues have implemented a practical spreadsheet/automated version called BARDI (Bayesian Adverse Reaction Diagnostic Instrument) and have now applied it to a number of practical adverse event problems.^{44,53,54}

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 32.3. Estimation of this overall probability, the “posterior probability,” is based on two components:

1. what is known prior to the event, the “prior probability,” which is based on clinical trial and epidemiologic data, and
2. what the likelihoods are or are not for drug causation of the components of the specific case, including its history, timing, characteristics, dechallenge and its timing components, rechallenge, and any other factors, such as multiple rechallenges.

The full application of this method requires knowledge of the clinical event, its epidemiology, and relatively specific information about the event’s characteristics and kinetics over time. Examples have been published for several types of event, including Stevens–Johnson Syndrome, renal toxicity, lithium dermatitis, and ampicillin associated colitis, agranulocytosis, and Guillain-Barré syndrome.^{23,44,53} Thus far, this approach appears to be useful for the analysis of the perplexing first events in new drug clinical trials,

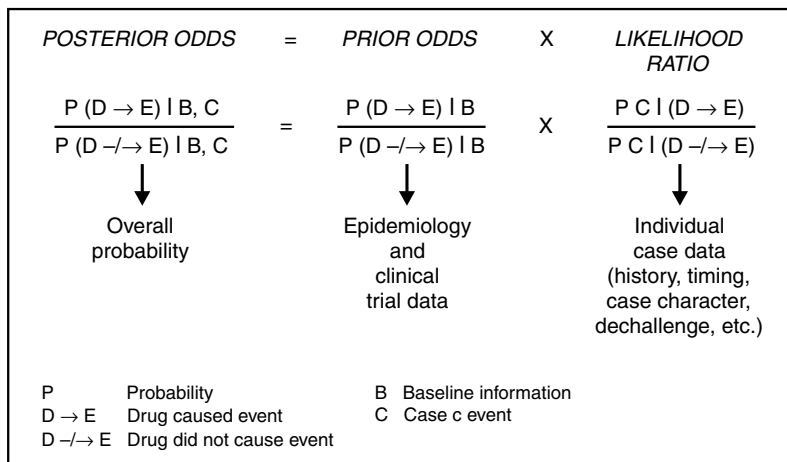


Figure 32.3. The basic equations for the Bayesian analysis of suspected drug associated events. These provide a structured, yet flexible and explicit approach to estimating the probability that an event is associated with one, or more, drugs, as described in the text and extensive literature dating from Auriche,²³ Lane,²⁵ and others. Since the prior probability estimate is dependent on explicit data from clinical trials and epidemiologic studies, this approach can provide a framework for specific event-related questions in these studies.

serious spontaneous adverse reaction reports, and possibly rare events discovered in both case-control and cohort pharmacoepidemiology studies, when standard methods of statistical analysis will not provide sufficient clues as to causality because of inadequate sample size.

With the logistic problem of the length of time required for the actual calculations minimized by automation, the major impediment to more general application of the Bayesian method is the state of the information required for robust analyses of events. There is no abundance of data on the incidence of most events and their occurrence in the presence and absence of most drugs (the required information for the prior probability). There are even fewer data available on the historical risk factors, the time course, and specific characteristics of the drug associated conditions, as opposed to the naturally occurring conditions. Although this lack of information is a current limitation, it represents both an important challenge and a framework for structuring further understanding. Benichou and collaborators have delved further into a mapping process of reactions by type in an attempt to begin classifications of specific drug associated disease, using acute liver disease as one model that incorporates qualitative clinical definitions of the disease into the judgement.^{58,59}

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 analysed separately, a sensitivity analysis of each information component can estimate its overall contribution to the final posterior odds or probability estimate. This, in turn, can be used to determine which information is pivotal. For example, if a tenfold difference in the estimate of the timing does not materially modify the overall posterior odds estimate, further efforts to determine the "best" estimate would not be worthwhile.
3. Because of the multistep approach to a judgement, 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 judgement. 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 analysis. For example, the timing and risk factor data on antibiotic associated diarrhea in three state Medicaid populations (Jones *et al.* unpublished) will provide extensive population data which can be used in one of the Bayesian analyses of this entity in single cases.^{25,51,52}

COMPARISON AMONG THE DIFFERENT METHODS FOR CAUSALITY ASSESSMENT

Several efforts have been made to evaluate and compare these methods. The 1983 conference in Crystal City involved the application of several of the methods to a standardized case, illustrating a considerable lack of concordance for some methods.¹⁸

A much more elegant and detailed evaluation of six representative algorithmic methods has been carried out more recently by Pere *et al.*,⁵⁵ who identified standard evaluation criteria and carried out an evaluation of 1134 adverse reactions using the various methods. Significantly, they found only moderate agreement between all pairs, and considerable disagreements on weightings of three

of the major criteria—timing, dechallenge, and alternate etiologies.

Given the current state of affairs, where a number of published methods exists, the choice of a method for use in evaluating individual adverse effects will likely be determined by a number of practical factors. These include the following.

1. *How the evaluation will be used.* This refers to both its short term use (e.g., a higher rating may be needed to result in a “signal” or need to report an event in a clinical trial) 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 judgement.* 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 judgement may be critical. Conversely, if little hinges upon the judgement, 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 judgements 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 (1, and see Chapter 8).
4. *The accrued value of thorough evaluations.* In some circumstances, the careful, rigorous evaluation of certain categories of drug associated event 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.⁵⁶ In some cases this results from data collection being focused to a much greater degree, as has been initiated in France by

Benichou *et al.*; here special reporting forms based on disease-specific criteria for events are being developed.^{56,57}

5. *Who will be carrying out the evaluation?* Although no specific studies have been carried out to evaluate the inter-rater differences among differently trained professionals, it is likely that the body of information held by each reviewer will have considerable impact on any of the methods used, including the Bayesian method.

THE FUTURE

The field of adverse reaction causality assessment has many unresolved issues, both methodologic and practical, which have been described in the preceding sections. Although there was an original hope that there would be some basis for a consensus method,¹⁸ the current state of the field would suggest that this is not likely to be the case, as again evidenced at the Third International APWI meeting in Paris (November 1992). 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.¹⁴ 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. Now that this barrier has been lifted, and with further use for practical applications, its potential may be realized. It is likely that the complex appearance of the algorithm of Kramer *et al.*⁸ likewise discourages its use in some sectors, although this has not been documented. Again, this is diminished with automation. Third, concern about the misuse of judgment terms or scores within the legal arena (see Chapter 9) has generated concern (34), particularly given the fact that there is no gold standard method.

All of these factors suggest the need for considerable further work. This work would appear to fall into at least three 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 reaction. This has long been recognized^{10, 18, 55} and is being implemented in some centers (e.g., France, University of Toronto). Further work in this area can have a major impact on:
 - (a) the collection of better information, using data collection instruments tailored to the event of interest, and
 - (b) the better definition of the dynamics and, ultimately, the mechanisms of certain types of drug induced condition.
3. Gathering of data on these *critical elements* in the course of both clinical trials and epidemiologic studies. Risk factor, history, timing, characteristics, and resolution patterns of adverse events should be described in these studies and incorporated into general data resources on the characteristics of medical events and diseases.
4. Further work on *automation* of the causality evaluation process. Global introspection is still widely used because of the cumbersome nature of many of the more complete methods. Fortunately, several methods are now being automated, including the French method,⁵⁵ the Venulet method (J. Venulet, personal communication), and the Bayesian BARDI method.⁵⁴ Convenient access to the proper questions, arrayed in logical order, as well as background data on the state of information to date, has the potential for radically changing the state of adverse reaction causality evaluation.

REFERENCES

1. Jones JK. Uses of drug-event assessment. *Drug Inf J* 1984; **18**: 233-40.
2. US Code of Federal Regulations 1987; **21**: 312.22 on Investigational New Drug procedures.
3. Feinstein AR. Clinical biostatistics. XLVII. Scientific standards vs. statistical associations and biologic logic in the analysis of causation. *Clin Pharmacol Ther* 1979; **25**: 481-92.
4. Hutchinson, TA, Lane DA. Assessing methods for causality assessment. *J Clin Epidemiol* 1989; **42**: 5-16.
5. Irey NS. Diagnostic problems in drug-induced diseases. In: Meyler L, Peck HM, eds, *Drug-Induced Diseases*, Vol. IV. Amsterdam: Elsevier, 1972; 1-24.
6. Irey NS. Adverse drug reactions and death: a review of 827 cases. *J Am Med Assoc* 1976; **236**: 575-8.
7. Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. *Clin Pharmacol Ther* 1977; **21**: 247-54.
8. Kramer MS, Leventhal JM, Hutchinson TA *et al*. An algorithm for the operational assessment of adverse drug reactions. I. Background, description, and instructions for use. *J Am Med Assoc* 1979; **242**: 623-32.
9. Naranjo CA, Busto U, Sellers EM, *et al*. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; **30**: 239-45.
10. Begaud B, Evreux JC, Jouglard J *et al*. Unexpected or toxic drug reaction assessment (imputation). Actualization of the method used in France. *Therapie* 1985; **40**: 111-8.
11. Emanuelli A, Sacchetti G. An algorithm for the classification of untoward events in large scale clinical trials. *Agents Actions* 1980; **7**: 318-22.
12. Jones JK. Adverse reactions in the community health setting: Approaches to recognizing, counseling, and reporting. *Family Community Health* 1985; **5**: 58-67.
13. Turner WM. The Food and Drug Administration algorithm. *Drug Inf J* 1984; **18**: 259-66.
14. Venulet J, Ciucci AG, Berneker GC. Updating of a method for causality assessment of adverse drug reactions. *Int J Clin Pharmacol Ther Toxicol* 1986; **24**: 559-68.
15. Castle WM. Assessment of causality in industrial settings. *Drug Inf J* 1984; **18**: 297-302.
16. Venulet J, Berneker GC, Ciucci AG, eds. *Assessing Causes of Adverse Drug Reactions*, London: Academic, 1982.
17. Herman RL, ed. *Drug-Event Associations: Perspectives, Methods, and Users*, proceedings of Drug Information Association Workshop, Arlington, VA, 1983.
18. Jones JK. Drug-event associations: a view of the current status. Epilogue. *Drug Inf J* 1984; **18**: 233-40.

19. Jones JK, Herman RL. Assessing adverse drug experiences [letter]. *Ann Intern Med* 1988; **108**: 160.
20. Venulet, J. APWI: past and present. *Drug Inf J* 1991; **25**: 229–33.
21. Lane DA. A probabilist's view of causality assessment. *Drug Inf J* 1984; **18**: 323–30.
22. de Finetti B. *Theory of Probability*. New York: Wiley, 1974.
23. Jones JK, Herman RL (guest eds). *Introduction, The Future of Adverse Drug Reaction Diagnosis: Computers, Clinical Judgement and the Logic of Uncertainty*, proceedings of the Drug Information Association Workshop, Arlington, VA, 1986. *Drug Inf J* 1986; **20**: 383.
24. Lane D. Causality assessment for adverse drug reactions: a probabilistic approach. In: Berry D, ed., *Statistical Methodology in the Pharmaceutical Sciences*. New York: Dekker, 1990; 475–507.
25. Lane DA, Kramer MS, Hutchinson TA *et al*. The causality assessment of adverse drug reactions using a Bayesian approach. *Pharm Med* 1987; **2**: 265–83.
26. Jones JK. A Bayesian approach to causality assessment. *Psychopharm Bull* 1987; **23**: 395–9.
27. CPMP Working Party on Pharmacovigilance. *Procedure for Causality Classification in Pharmacovigilance in the European Community*, EC Document III/3445/91-EN, final, 1991.
28. Jones JK. Causality assessment of suspected adverse drug reactions: a transatlantic view. *Pharmacoepidemiol Drug Safety* 1992; **1**: 251–60.
29. Miremont G, Haramburu F, Begaud B, Pere JC, Dangoumou J. Adverse drug reactions: physicians' opinions versus a causality assessment method. *Eur J Clin Pharmacol* 1994; **46**: 285–9.
30. Castle W, Baker A. Communication of ADR information: the viewpoint of a British company. *Drug Inf J* 1985; **19**: 375–80.
31. Stephens M. Assessment of causality in industrial settings. *Drug Inf J* 1984; **18**: 307–14.
32. Stephens M. *The Detection of New Adverse Drug Reactions*. London: Plenum, 1985.
33. Auriche M. Approche bayesienne de l'imputabilité des phénomènes indésirables aux médicaments. *Thérapie* 1985; **40**: 301–6.
34. Freilich WB. Legal perspectives in causality assessment. *Drug Inf J* 1984; **18**: 211–7.
35. Margulies, JB. Epidemiologic causation in the courtroom: square pegs in round holes? *Drug Inf J* 1991; **25**: 217–27.
36. Henderson TW. Legal aspects of disease clusters. Toxic tort litigation: medical and scientific principles in causation. *Am J Epidemiol* 1990; **132**: S69–78.
37. Mashford ML. The Australian method of drug–event assessment. *Drug Inf J* 1984; **18**: 271–3.
38. Wiholm BE. The Swedish drug–event assessment methods. *Drug Inf J* 1984; **18**: 267–9.
39. Dangoumau J, Begaud B, Boisseau A, *et al*. Les effets indésirables des médicaments. *Presse Med* 1980; **9**: 1607–9.
40. Editorial. Improving reports of adverse drug reactions. *Br Med J* 1984; **289**: 898–9.
41. Venulet J. Incomplete information as a limiting factor in causality assessment of adverse drug reactions and its practical consequences. *Drug Inf J* 1986; **20**: 423–31.
42. 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.
43. Louik C, Lacouture PG, Mitchell AA *et al*. A study of adverse reaction algorithms in a drug surveillance program. *Clin Pharmacol Ther* 1985; **38**: 183–7.
44. Naranjo, CA, Lanctot, KL, Lane DA. The Bayesian differential diagnosis of neutropenia and antiarrhythmic agents. *J Clin Pharmacol* 1990; **30**: 1120–7.
45. Jones JK. Epidemiologic perspective on causality assessment for drug associated events. *Drug Inf J* 1986; **20**: 413–22.
46. Kahneman D, Slovic P, Tversky A. *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 1982.
47. Koch-Weser J, Sellers EM, Zacest R. The ambiguity of adverse drug reactions. *Eur J Clin Pharmacol* 1977; **11**: 75–8.
48. Karch FE, Smith CL, Kernzer B *et al*. Adverse drug reactions—a matter of opinion. *Clin Pharmacol Ther* 1976; **19**: 489–92.
49. Kramer MS. Assessing causality of adverse reactions: global introspection and its limitations. *Drug Inf J* 1986; **20**: 433–7.
50. Herman RL, Fourrier A. *List of References: Third International APWI Meeting*. Paris, 1992.
51. Hutchinson TA, Dawid AP, Spiegelhalter DJ, Cowell RG, Roden, S. Computerized aids for probabilistic assessment of drug safety I: a spreadsheet program. *Drug Inf J* 1991; **25**: 29–39.
52. Hutchinson TA, Dawid AP, Spiegelhalter DJ, Cowell RG, Roden, S. Computerized aids for probabilistic assessment of drug safety I: an expert system. *Drug Inf J* 1991; **25**: 41–8.
53. Lanctot KL, Naranjo, CA. Using microcomputers to simplify the Bayesian causality assessment of adverse drug reactions. *Pharm Med* 1990; **4**: 185–95.
54. Naranjo CA, Lanctot, KL. Microcomputer assisted Bayesian differential diagnosis of severe adverse drug reactions to new drugs: a 4-year experience. *Drug Inf J* 1991; **25**: 243–50.
55. Pere JC, Begaud B, Haramburu F, *et al*. Computerized comparison of six adverse drug reaction assessment procedures. *Clin Pharmacol Ther* 1986; **40**: 451–61.
56. Benichou C. *Guide Pratique de Pharmacovigilance*. Paris: Pradel, 1992.

57. Benichou C, Danon G. Experts' opinion in causality assessment: results of consensus meetings. *Drug Inf J* 1991; **25**: 251–5.
58. Benichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs—I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. *J Clin Epidemiol* 1993; **46**: 1323–30.
59. Benichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs—II. An original model for validation of drug causality assessment methods: case reports with positive rechallenge. *J Clin Epidemiol* 1993; **46**: 1331–6.

33

The Use of Randomized Controlled Trials for Pharmacoepidemiology Studies

SAMUEL M. LESKO

Slone Epidemiology Unit, Schools of Public Health and Medicine, Boston University, Brookline, MA, USA

ALLEN A. MITCHELL

Slone Epidemiology Unit, Schools of Public Health and Medicine, Boston University, Brookline, MA, USA

INTRODUCTION

When properly conducted, randomized controlled trials (RCTs) are considered the gold standard for demonstrating the effectiveness of a new medication because they provide unbiased estimates (see Chapter 2). While generally used for studies of beneficial drug effects (see also Chapter 34), too often pharmacoepidemiologists overlook the fact that 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 on aggregate include at most a few thousand patients. These studies are sufficiently large to provide evidence of a beneficial clinical effect and exclude large increases in risk of common adverse clinical events. However, premarketing trials are

rarely large enough to detect small differences in the risk of common adverse events or to reliably estimate the risk of rare events, whether serious or trivial. Identification and quantification of these potentially important risks require large studies, which typically are conducted after a drug has been marketed. Because of design complexity and costs, large controlled trials have not generally been considered in the pharmacoepidemiologist's armamentarium for the postmarketing evaluation of drugs. Until recently, we too did not consider this approach for our postmarketing studies. However, our search for the best method to assess the risk of serious but rare adverse reactions to pediatric ibuprofen caused us to expand our view.¹ The experience that led to that change serves as the basis for this chapter and may prompt others to consider randomized trials for the postmarketing assessment of drug safety.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY 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. In this chapter, we will consider the role of postmarketing randomized trials in assessing only the risks of medications; however, the same principles can be applied to the postmarketing evaluation of the benefits of medications (see also Chapter 34).

As noted above, premarketing studies are typically too small to detect modest differences in the incidence rates (e.g., relative risks of 2.0 or less) for common adverse events or even large differences in the incidence rates for rare events, such as those that affect one per 1000 treated patients. 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. For example, following the introduction of angiotensin converting enzyme (ACE) inhibitor use in patients with congestive heart failure, case reports of severe hypotension began to appear in the literature. Although similar events were noted after initial use of other medications (e.g., vasodilators) in congestive heart failure patients, reliable estimates of the risk for different classes of medication were not available. Because of the high prevalence of the indication, differences in risk too small to be detected by conventional RCTs were judged to be clinically important, and a large RCT was conducted to resolve the question.² Modest risks are especially relevant to drugs being considered for over-the-counter (OTC) sale, because these agents are likely to be very commonly used and are widely viewed by the public as safe. After a drug has been licensed, large observational studies are typically used to satisfy the sample sizes needed to identify (or rule out) such risks. The respective strengths and weaknesses of these designs are discussed elsewhere in this volume (see Chapter 2). However, potential confounding is a major con-

cern for virtually every observational study. Uncontrolled or incompletely controlled confounding can easily account for modest associations between a drug and an adverse clinical event. For example, in the relation between phenylpropanolamine and cerebrovascular disease, obesity increases both the likelihood of exposure to the drug and the risk of a cerebrovascular accident; thus, body weight must be controlled in any analysis of this association. The challenge to the pharmacoepidemiologist is to recognize those factors that represent potential confounders and then control for their effects. To do so requires the relevant information to be included in the data to be analyzed, but information on important confounding factors is frequently incomplete or unavailable. Although surrogate variables are often used (e.g., years of education to reflect socioeconomic status), these may be poor measures of the underlying confounding factor, and their control therefore may not eliminate confounding.

An investigator observing a crude association between a drug and an effect attempts to control for confounding by adjusting for one or more factors. If a crude odds ratio (or relative risk) of 5.0 (for example) remains essentially unchanged after controlling for all known confounders, residual confounding is usually not considered an important concern; although the true (unconfounded) odds ratio may be somewhat smaller than the unadjusted estimate, it is generally assumed to be of similar magnitude. On the other hand, if the adjusted odds ratio (or relative risk) is closer to the null value of 1.0, there is empirical evidence of confounding in the data, and the adjusted odds ratio is usually considered the “best” (least biased) measure of the association. However, it is not possible to determine whether there remains any residual confounding in this best estimate, which if *completely* controlled might reveal that the true association is still weaker or even nonexistent.

We have direct experience with this concern. Infants treated in newborn intensive care units often receive medications and intravenous fluids through indwelling catheters, and low doses of heparin are periodically infused to help maintain the patency of

these catheters. In 1986, we published the results of a case-control study of the use of intravenous heparin in relation to the risk of intraventricular hemorrhage (IVH) in low birth weight infants.³ We compared 66 infants with IVH (cases) to 254 infants with no evidence of IVH (controls), matched on study hospital and duration of observation. Compared to no heparin exposure, the matched odds ratio for heparin exposure on the day prior to detection of IVH was 14 (95% confidence interval (CI), 5.4–34). As additional potential confounders were taken into account, the magnitude of the association became progressively smaller (Table 33.1). Adjustment for the matching factors, birth weight, volume of parenteral fluids administered, and the presence of pneumothorax by logistic regression, reduced the odds ratio to 3.9 (95% CI, 1.4–11), which did not change further when other potential confounding factors were added to the multivariate model. Although we described an observation that was statistically significant, biologically plausible, and clinically important, we concluded that control of confounding may have been incomplete and that, “... the association could have been partly, or even wholly, due to the severity of the infants’ underlying conditions rather than to the use of heparin.” We suggested that the question could only be answered by a randomized trial. A second observational study also found an increased risk (odds ratio = 1.96) among infants who received doses of heparin above the lowest quartile of exposure.⁴ Uncertainty about the association persisted until 1997, when results were published from a randomized, double blind trial of heparin added to umbilical catheters used to treat premature infants.⁵ In this study involving 113 infants, Chang

et al. found no difference in the incidence of intraventricular hemorrhage between the heparin treated and control groups ($p = 0.6$). Although the odds ratios from the earlier observational studies were moderately large (3.9 and 1.96) and statistically significant, these “best” estimates of risk were likely due to confounding by one or more factors not completely controlled in the analyses. As this example demonstrates, when residual confounding is a possible explanation for an apparent association, additional research is needed before a change in clinical practice is warranted.

Weak associations deserve particular attention. Although there are important exceptions, the general view is that the stronger the association, the more likely the observed relationship is causal. This is not to say that a weak association (e.g., a relative risk ≤ 1.5) can never be causal; rather, it is more difficult to be certain of it because such associations, even if statistically significant, can easily be an artifact of confounding. As an example, consider an analysis where socioeconomic status is a potential confounder and education is used as a surrogate for this factor. Because the relation between years of education completed (the surrogate) and socioeconomic status (the potential confounder) is, at best, imperfect, analyses controlling for years of education can only partially control for confounding. This leads to the familiar caveat in reports of observational studies, “...residual confounding may account for the observed association.” This qualification is no more appropriate than for studies reporting weak associations. As a consequence, even after rigorous efforts have been made to control for confounding, some seasoned

Table 33.1. Effect of potential confounding factors on the relation between heparin exposure and intraventricular hemorrhage in 320 premature infants

Model	Potential confounders included	Odds ratio ^a	95% confidence interval
1	Hospital and duration of observation	14	5.4–34
2	Model 1 + birth weight	7.5	2.8–20
3	Model 2 + fluids	4.4	1.6–12
4	Model 3 + pneumothorax	3.9	1.4–11

^a Calculated by logistic regression controlling for the potential confounders listed. Adapted from reference 3.

epidemiologists consider small relative risk estimates to be most compatible with no association, regardless of the confidence interval (or p value). Whether or not one subscribes to this view, it is advisable to use extreme caution in making causal inferences from small relative risks derived from observational studies.

When there is concern about residual confounding prior to embarking on an observational study, one may wish to consider using a nonobservational study design. We faced just this situation when we considered how to best assess the safety of pediatric ibuprofen. Ibuprofen is a nonsteroidal anti-inflammatory drug (NSAID) that has become widely used among adults in the US, first by prescription and then as an OTC drug. In 1989, ibuprofen suspension was licensed as a prescription product for fever control in children. It was approved based on premarketing studies in children, which established that it was effective and safe for use under a doctor's supervision. Events known to occur in adults using ibuprofen, such as acute gastrointestinal bleeding, acute renal failure, and anaphylaxis, were either not observed at all during the premarketing trials in children or occurred so infrequently that it was not possible to obtain reliable estimates of the risk. In addition, it was at least theoretically possible that Reye's syndrome (a toxic encephalopathy in children associated with another NSAID, aspirin) might be associated with ibuprofen use in children. Other events, possibly unique to children, might also be associated with this drug. Thus, premarketing studies were unable to exclude even a substantially increased risk of important rare but serious adverse reactions.

Once available OTC, pediatric ibuprofen would likely be widely used for the treatment of fever, which is typically a minor and self-limited condition. (We will not discuss the debate over when it is appropriate to treat fever in children.) Given the generally benign nature of this indication, it is reasonable to require greater assurance of safety than may be expected for a drug used to treat a life threatening illness. Further, an effective antipyretic with an excellent record of safety in children, acetaminophen, had been available OTC in the US for more than 20 years. For these reasons, the US

Food and Drug Administration required additional data concerning the risk of rare but serious adverse events before it would approve pediatric ibuprofen for OTC sale.

What approach would best provide this information? Observational postmarketing studies, especially case-control studies, are one source of data for very rare conditions. However, the circumstances surrounding ibuprofen use in 1989-90 raised serious concern about whether observational studies could adequately control confounding. Specifically, prior to the availability of pediatric ibuprofen, febrile children in the US received no antipyretic or were given acetaminophen, which was generally considered safe by both physicians and parents. On the other hand, because ibuprofen was available only by prescription, treatment with this drug required contact with a physician. In addition, for fever less than 102.5 °F, the recommended dose of prescription ibuprofen is 5 mg kg⁻¹, whereas for fever of 102.5 °F or greater, the dose is 10 mg kg⁻¹. Both its status as a prescription medication and the two-tier dosing schedule predicted that ibuprofen would be used for more severe illness than acetaminophen. This prediction was supported by a survey of 108 physicians (61 pediatricians, 47 family practitioners) conducted in 1992.¹ More than half of the physicians in the study reported that they treated children with ibuprofen after acetaminophen failed, but none reported using acetaminophen only when ibuprofen was not effective. Further, both the minimum age and temperature at which the physicians recommended using these drugs were higher for ibuprofen than acetaminophen. It seemed clear that pediatric ibuprofen would be most commonly used among children whose illness was relatively severe, or whose fever was particularly high or unresponsive to acetaminophen. Because of the greater severity of illness (and potential exposure to antibiotics or other medications), there was a reasonable basis to believe that ibuprofen users would experience relatively high rates of adverse clinical events, *unrelated to the ibuprofen itself*. It was apparent, then, that to provide a valid assessment of the risks of pediatric ibuprofen, a study must be able to distinguish the risks of the drug from the risks

associated with the illness for which ibuprofen was given.

METHODOLOGIC PROBLEMS TO BE SOLVED BY PHARMACOEPIDEMIOLOGY RESEARCH

The phenomenon described above for pediatric ibuprofen is known as confounding by indication (also referred to as indication bias, channeling, confounding by severity, or contraindication bias). According to Slone *et al.* confounding by indication exists when “patients who receive different treatments...differ in their risk of adverse outcomes, independent of the treatment received.”⁶ In general, confounding by indication occurs when an observed association between a drug and an outcome is due to the underlying illness (or its severity) and not to any effect of the drug. Put another way, confounding by indication occurs when the risk of an adverse event is related to the *indication* for medication use but not the use of medication itself. As with any other form of confounding, one can, in theory, control for its effects if one can reliably measure the severity of the underlying illness. In practice, however, this is not easily done. Confounding by indication is a particular concern in a number of settings (see Chapters 34 and 43). When there is a single therapy for an illness, and all patients receive the therapy (i.e., are “channeled” to the treatment), it is not possible to control for confounding in an observational study simply because no patients are left untreated to serve as controls. For example, it is standard practice to administer artificial surfactant to premature infants at risk for respiratory distress syndrome of the newborn. If any infants are not treated, they are likely to differ from treated infants in that they may have a very mild form of the illness, or they may have a major congenital malformation and not be expected to survive. Thus, they are also likely to have different risks for many clinical outcomes. While it may be rare for all patients with a given illness to be treated in exactly the same way, this situation is not unusual for subgroups of patients. For example, all patients with diabetes are not treated

with insulin, but patients with type I (insulin-dependent) diabetes are. In general, observational studies are most informative when patients receiving different medications are similar. Cohort studies will be compromised if there is no reasonable alternative to the study treatment, including no treatment, to serve as a control. Case-control studies may be infeasible if one cannot identify controls that, aside from any effect of the exposure, are equally at risk of having the outcome diagnosed as the cases.

When there is at least one alternate treatment option and it is possible to control for obvious confounding, observational studies can contribute to our understanding of a medication’s risks, particularly where the adjusted relative risk is large. However, as discussed above, a small relative risk (e.g., 1.3) can easily be an artifact of confounding by an unknown factor or by incomplete control of a recognized confounder.

When confronted with the task of assessing the safety of a marketed drug product, the pharmacoepidemiologist must evaluate the specific hypothesis to be tested and estimate the magnitude of the hypothesized association and determine whether confounding by indication is possible. If incomplete control of confounding is likely, it is important to recognize the limitations of observational research designs and consider conducting an RCT. There is nothing inherent in an RCT that precludes a pharmacoepidemiologist from designing and carrying out these studies. To the contrary, the special skills of a pharmacoepidemiologist can be very useful, in performing large scale RCTs after a drug is marketed.

OVERVIEW OF CLASSIC RCTS

In the evaluation of drugs, RCTs are most commonly used during the premarketing phases of drug development to demonstrate a drug’s efficacy (and to gather general information concerning safety). By randomization, one hopes to make the distributions of confounding factors (both known and unknown) equal in all groups. If the study is sufficiently large, the assigned treatment is the most likely explanation for any

observed differences in the clinical outcomes (improvement in the illness or the occurrence of adverse clinical events) between the treatment groups. By definition, participants in observational studies are not assigned treatment at random. As we have seen, the choice of treatment may be determined by the stage or severity of the illness or by the patient's poor response to or adverse experience with alternative therapies, which can introduce bias.

Sample Size

In homogeneous populations, balanced treatment groups can be achieved with relatively small study sizes. In heterogeneous populations (e.g., all children less than 12 years of age), a large sample size may be required to insure the equal distribution of uncommon confounders between study groups. Study size is determined by the need to assure balance between treatment groups and the magnitude of the effect to be detected. Large studies minimize the chance that the treatment groups are different with respect to potential confounders and permit the detection of small differences in clinical outcomes.

Blinding

Blinding is used to minimize detection bias, and is particularly important where the outcome is at all subjective. Reporting of subjective symptoms by study participants and the detection of even objectively defined outcome events may be influenced by knowledge of the medications the patient is using. For example, if a patient complains of abdominal pain, a physician may be more likely to perform a test for occult blood in the stool if the patient was being treated with ibuprofen than acetaminophen. Thus, followup data collection will be unbiased if both parties are unaware of the treatment assigned. Blinding may not be possible for nondrug treatments such as diet, exercise, and surgery, and double blinding may be difficult to achieve and maintain in drug studies as well, particularly if either the study or control medication produces specific symptoms (i.e., side-effects)

or easily observable physiologic effects (e.g., change in pulse rate or blood pressure).

Choice of Control Treatment

The hypothesis being tested determines the choice of control treatment. Placebo controls are most useful for making comparisons with untreated disease but may not represent standard of care and have been challenged as unethical.⁷ Further, it may be difficult to maintain blinding in placebo controlled studies, as noted above. Studies employing an active control typically utilize common drug treatments, which frequently represent the standard of care. Although often considered more ethical and easier to keep blinded because the illness and symptoms are not left untreated, these studies do not permit comparison with the natural history of the illness.

Data Collection

Data collection in 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 intervals during followup. In addition to the data needed to test the hypothesis of a clinical benefit, premarketing trials of medications must also assess safety and therefore must collect extensive data on symptoms, physical signs, and laboratory evaluations. Such data collection contributes substantially to the high cost of these trials.

Data Analysis

In observational studies, data analyses are driven by the hypothesis being tested and may be quite complex. In contrast, analysis of the primary hypothesis in many clinical trials is straightforward and involves a comparison of some measure of the outcome event (which may be either a continuous or categorical variable) in different groups. While analyses involving repeated measures, subgroups of study subjects, or adjustment to control for incomplete or ineffective randomization may be performed, they add complexity.

Methodologic strengths notwithstanding, there are several features of the RCT that limit its use as a postmarketing study design. First, the complexity and cost of traditional premarket RCTs, with their detailed observations and complex followup, make such very large studies generally infeasible. Second, it may be unethical to conduct a study in which patients are randomly assigned a potentially harmful treatment, such as an RCT to test the hypothesis that cigarette smoking increases the risk of heart disease. However, if the study can be simplified and use the epidemiologist's tools to track patients and collect followup data, it may be possible to both control costs and make a study feasible. The ethical dilemma can be resolved by studying only questions that are truly important to the public's health and for which the answers are not known.

Generalizability of Results

The usual clinical trial conducted during the premarketing evaluation of a drug almost always involves highly selected patients; as a consequence, the results of the trial may not be generalizable to the large numbers of patients who may use the medication once it is licensed. Pharmacoepidemiologists may be more attracted to observational studies because 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.

CURRENTLY AVAILABLE SOLUTIONS

LARGE SIMPLE TRIALS

Large, simple trials (LSTs) may be the best solution when it is not possible to completely control confounding by means other than randomization. If the volume and complexity of data collection can be kept to a minimum, there is no reason that large trials cannot be conducted. Indeed, beginning with the Salk vaccine trial, LSTs have been used to test the efficacy of medical interventions.⁸⁻¹⁹ This approach has also been used successfully to evaluate the risk of adverse

drug effects when the more common observational designs have been judged inadequate.^{2,20} These studies are really just very large randomized trials made simple by reducing data collection to the minimum needed to test only a single hypothesis (or at most a few hypotheses). Randomization of treatment assignment is the key feature of the design, which controls for confounding by known and unknown factors. The large study size provides the power needed to evaluate small risks, either absolute or relative.

How Simple is Simple?

Yusuf *et al.* have suggested that very large randomized studies of treatment related mortality need collect only data concerning the vital status of participants at the conclusion of the study.²¹ Because the question of drug safety frequently concerns outcomes less severe than mortality, these ultrasimple trials may not be sufficient. Hasford has suggested a somewhat less restrictive approach to data collection, in which "large trials with lean protocols" include only *relevant* baseline, followup, and outcome data.²² Collecting far fewer data than is common in the usual RCT is the key feature of both approaches. With simplified protocols that take advantage of epidemiologic followup methods, very large trials can be conducted to test hypotheses of interest to pharmacoepidemiologists.

Power/Sample Size

Study power is not simply a function of the number of subjects enrolled. It is related to the number of events observed during the course of the study, which in turn is a function of the incidence rate for the event, the sample size, and the duration of observation (or followup). Power requirements can be satisfied by studying a population at high risk, enrolling a large sample size, or conducting followup for a prolonged period. The appropriate approach will be determined by considering the goal of the study and the hypothesis to be tested. Allergic or idiosyncratic events require a very large study population, and events with long latency periods may be best studied with long duration

followup. 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 be inappropriate to assess the risk of these events in younger adults or children.

Data Elements

The data collection process can be kept simple by restricting the study to a few primary endpoints that satisfy the study hypothesis, are objective, are easily identified, and are verifiable. Epidemiologists may need to overcome their predisposition to data collection when it comes to secondary outcomes (i.e., those that do not directly relate to the study hypothesis), as these must be ignored to eliminate unnecessary data collection. Because confounding is controlled by randomization, data on potential confounders need not be collected. A few basic demographic variables can be collected in order to confirm that effective randomization was achieved.

Data Collection

The data collection process itself can be streamlined to keep the study simple. Followup data can be collected by mailed questionnaires or telephone interviews conducted directly with the study participants. Because the study will be limited to clear and objective outcomes, which can be confirmed by medical record review or other means, self-report by the study participants can be an appropriate source of followup data.

The primary advantage of this simplicity is that it allows very large groups of study participants to be followed at reasonable cost. The tradeoff, of course, is that a simple trial cannot answer all possible questions about the safety of a drug but must be limited to testing, at most, a few related hypotheses.

WHEN IS A LARGE SIMPLIFIED RANDOMIZED TRIAL APPROPRIATE?

LSTs are appropriate when all of the conditions in Table 33.2 apply.

Table 33.2 Conditions appropriate for the conduct of a large randomized trial

-
1. The research question is important.
 2. Genuine uncertainty exists about the likely results.
 - 3a. When the absolute risk is small and confounding by indication is likely
- or
- 3b. The relative risk is small, regardless of the absolute risk.
 4. Important effect modification (interaction) is unlikely.
-

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 does 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.” This was originally described by Gray *et al.* as a simple criterion to assess subject eligibility in LSTs.²³ It states that “... both patient and doctor should be *substantially uncertain* about the appropriateness, for this particular patient, of each of the trial treatments. If the patient and doctor are *reasonably certain* that one or other treatment is inappropriate then it would not be ethical for the patient’s treatment to be chosen at random” (italic in the original). We support this principle and would extend its use to evaluate when it is appropriate to conduct an LST to test a given hypothesis related to the risk of an adverse clinical event. Very large randomized trials are justified only when there is true uncertainty about the risk of the treatment in the population.

Apart from considerations of benefit, it would not be ethical to subject large numbers of patients to a treatment that was reasonably believed to place them at increased risk, however small, of a potentially serious or permanent adverse clinical event. The concept of uncertainty can thus be extended to include a global assessment of the combined risks and benefits of the treatments being compared. One treatment may be known to provide therapeutic benefits that are superior to an alternative, but it may be unknown whether the risks of important side-effects outweigh the therapeutic advantage. For example, the antiestrogen tamoxifen may improve breast cancer survival, but may do so only at the cost of an increased risk of endometrial cancer. Appropriately a randomized trial was undertaken to resolve uncertainty in this and similar situations.¹⁹

Power and Confounding

LSTs will only be needed if (a) the *absolute* risk of the study outcome is small and there are concerns about confounding by indication, or (b) the *relative* risk is small (in which case, there are inherent concerns about residual confounding from any source).²¹ 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 were 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.²¹ In other words, the available evidence should suggest that the association will be qualitatively similar in all patient subgroups. Variation among subgroups in the strength of the association is acceptable, but there should be no suggestion that the effect would be completely reversed in one or more subgroups. Because of the limited data

available in a truly simple trial, it may not be possible to test whether an interaction has occurred, and the data collected may not be sufficient to identify relevant subgroups. Because randomization only controls confounding for comparisons made between the groups that were randomized, subsets of these groups may not be strictly comparable with respect to one or more confounding factors. Thus, if clinically important interaction is considered likely, additional steps must be taken to permit the appropriate analyses (e.g., stratified randomization). This added complexity may result in a study that is no longer a truly simple trial.

WHEN IS AN LST FEASIBLE?

LSTs are feasible when all of the conditions in Table 33.3 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

Table 33.3 Conditions which make a large, simple randomized trial feasible

-
1. The study question can be expressed as a simple testable hypothesis.
 2. The treatment to be tested is simple (uncomplicated).
 3. The outcome is objectively defined (e.g., hospitalization, death).
 4. Epidemiologic followup methods are appropriate.
 5. A cooperative and motivated population is available for study.
-

with LSTs. They are likely to be commonly used, so that it will be easy to enroll large numbers of patients, and the results will be applicable to a large segment of the population. Complex therapeutic protocols are difficult to manage, reduce patient compliance, and by their very nature may not be compatible with the simple trial design.

Objective and Easily Measured Outcomes

The outcomes to be studied should be objective, easy to define ("simple"), and easy to recall. An example might include hospitalization for acute gastrointestinal bleeding. Study participants may not 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 followup methods (e.g., questionnaires, telephone interviews, or linkage with public vital status records). On the other hand, clinical outcomes which can be reliably detected only by detailed in-person interviews, physical examinations, or extensive physiologic testing are not as amenable for study in simple trials.

Cooperative Population

Particularly in LSTs, a cooperative and motivated study population 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.^{24,25} Because of the participants' knowledge of medical conditions and symptoms and participation in the US health care system, relatively sophisticated information could be obtained using mailed questionnaires, and even biologic samples could be collected. Success of the Boston University Fever Study was also largely due to parents whose motivation and cooperation were encour-

aged by their private physicians who had invited them to participate in the study.²⁰

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 in private practice or members of large health care organizations) can participate by recruiting eligible patients and obtaining informed consent. Alternative methods to identify and enroll eligible subjects (e.g., direct mailings to professional groups, print advertisements) may be appropriate for some studies. Because success depends on the cooperation of multiple health care providers and a large number of patients, it may be best to limit the extent of each individual's involvement.

To facilitate patient recruitment and to maximize generalizability of the results, minimal restrictions should be placed on patient eligibility. As Gray *et al.* have said, "Any obstacle to simplicity is an obstacle to large size, and the wider the range of the patients studied, the wider the generalizability of the results will be."²³ 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. 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. For studies requiring immediate treatment, patients can be registered and treatment assigned by telephone contact with the coordinating center.

Particularly in studies requiring a long duration of medication use, validity may be seriously compromised by poor compliance with the treatment regimen. A run-in period prior to randomization can be used to identify patients who are unable or unwilling to adhere to a chronic treatment regimen and are likely to drop out of the study. During the run-in period, eligible subjects are given a "test" medication and their compliance with the protocol is assessed. Patients who cannot comply with the protocol are withdrawn from the trial. Patients who remain in the study 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.²⁵ As a consequence, however, the data cannot be used to assess the risk of gastrointestinal bleeding following aspirin use.

Importance of Complete Followup

Because dropouts and losses to followup may not be random but rather may be related to treatment side-effects, it is important to make every effort to obtain followup data on all enrolled subjects. For example, a study that has followup data on 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. Beyond choosing a motivated and interested study population, the investigators can minimize losses to followup by maintaining regular contact with all study participants. Regular mailings of supplies of medication or a study newsletter can be helpful, and memory aids such as medication calendar packs or other devices may help maintain compliance with chronic treatment schedules. In

addition, followup data collection itself can help maintain contact with study participants.

Followup Data Collection

An important element of a successful LST is that followup data collection is the responsibility of the central study staff. Busy health care providers frequently cannot commit the time required to consistently obtain even minimal but specific followup data from large numbers of subjects. However, the clinician who originally enrolled the subject may be able to provide limited followup data (e.g., vital status) or a current address or telephone number for the occasional patient who would otherwise be lost to followup. A questionnaire delivered by mail, supplemented by telephone interviews when needed, has been shown to work quite well.²⁰ The response rate will likely be greatest if the questions are both simple and direct and the time required to complete the questionnaire is limited. 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 followup.

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 followup) may not be fully representative of the original randomization groups and may not be comparable with respect to confounders. Despite all efforts, complete follow-up is rarely achieved, and because only the original randomization groups can be assumed to be free of confounding, at least one analysis involving all enrolled study subjects (i.e., an intention-to-treat analysis) should be performed. Also, unless a stratified randomization scheme was used, one cannot be certain that unmeasured confounding variables will be evenly distributed in subgroups of participants, and the smaller the subgroup, the greater the potential for imbalance. Therefore, subgroup analyses 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 often appropriate to monitor the accumulating data over the course of the study. The study may sometimes be ended prematurely if participants experience unacceptable risks, if the hypothesis can be satisfactorily tested earlier than anticipated, or if it becomes clear that a statistically significant result cannot be achieved, even if the study were to be completed as

planned. A data monitoring committee, independent of the study investigators, can conduct periodic reviews of the data using an appropriate group sequential analysis procedure to preserve the study's overall type I error rate.^{26,27}

THE FUTURE

With accelerated approval of new medications and rapid increases in their use, we may see a greater need for large postmarketing studies capable of randomizing exposures in order to assess small differences in risk. In the absence of techniques that reliably control for confounding by indication in observational studies, there may be a growing need for LSTs to evaluate larger relative risks. 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. 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 followup data. It would still be necessary to identify eligible subjects, obtain consent, and randomize treatment. 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 study medication. Though 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, it would at least be possible to accurately estimate the absolute risk of important events, whatever their cause, among exposed subjects. An alternative approach would

be to randomize to different doses, when possible, and search for a dose–response relationship.

It is clear that very large simple controlled trials of drug safety can be successfully carried out.^{2,20} It is less clear, however, how frequently the factors that indicate the need for a very large trial (Table 33.2) will converge with those that permit such a trial to be carried out (Table 33.3). As pharmaco-epidemiologists become more comfortable with LSTs, we may see more of them being conducted, and new methods of subject recruitment and more efficient sources of followup data are likely to be developed.

REFERENCES

- Mitchell AA, Lesko SM. When a randomized controlled trial is needed to assess drug safety: the case of pediatric ibuprofen. *Drug Safety* 1995; **13**: 15–24.
- 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.
- Lesko S, Mitchell AA, Epstein MF, Louik C, Giacoia GP, Shapiro S. Heparin use as a risk factor for intraventricular hemorrhage in low-birth-weight infants. *New Engl J Med* 1986; **314**: 1156–60.
- Malloy MH, Cutter GR. The association of heparin exposure with intraventricular hemorrhage among very low birth weight infants. *J Perinatol* 1995; **15**: 185–91.
- Chang GY, Leuder FL, DiMichele DM, Radkowski MA, McWilliams LJ, Jansen RD. Heparin and the risk of intraventricular hemorrhage in premature infants. *J Pediatr* 1997; **131**: 362–6.
- Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. A policy perspective. *Ann Intern Med* 1979; **90**: 257–61.
- Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *New Engl J Med* 1994; **331**: 394–8.
- Francis T Jr, Korn R, Voight R, Boisen M, Hemphill F, Napier J, Tolchinsky E. An evaluation of the 1954 poliomyelitis vaccine trials: summary report. *Am J Public Health* 1955; **45** (suppl): 1–50.
- 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; **2**: 57–66.
- ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988; **2**: 349–60.
- ISIS-3 (Third International Study of Infarct Survival) Collaborative Group. A randomised trial of streptokinase vs tissue plasminogen activator vs antistreptolysin and of aspirin plus heparin vs aspirin alone among 41,299 cases of suspected acute myocardial infarction. *Lancet* 1992; **339**: 753–70.
- ISIS-4 Collaborative Group. Fourth International Study of Infarct Survival: Protocol for a large, simple study of the effects of oral mononitrate, of oral captopril, and of intravenous magnesium. *Am J Cardiol* 1991; **68**: 87D–100D.
- Gruppo Italiano per lo Studio della Streptochinasi nell'infarto miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986; **1**: 397–402.
- Gruppo Italiano per lo Studio della Streptochinasi nell'infarto miocardico (GISSI). GISSI-2: A factorial randomized trial of alteplase versus streptokinase and heparin versus no heparin among 12490 patients with acute myocardial infarction. *Lancet* 1990; **336**: 65–71.
- Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *New Engl J Med* 1989; **321**: 129–35.
- GUSTO investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *New Engl J Med* 1993; **329**: 673–82.
- Yusuf S. Reduced mortality and morbidity with the use of angiotensin-converting enzyme inhibitors in patients with left ventricular dysfunction and congestive heart failure. *Herz* 1993; **18** (suppl 1): 444–8.
- Gerstein HC, Bosch J, Pogue J, Taylor DW, Zinman B, Yusuf S. Rationale and design of a large study to evaluate the renal and cardiovascular effects of an ACE inhibitor and vitamin E in high-risk patients with diabetes. The MICRO-HOPE Study. Microalbuminuria, cardiovascular, and renal outcomes. Heart Outcomes Prevention Evaluation. *Diabetes Care* 1996; **19**: 1225–8.
- Fisher B, Costantini JP, Wickerham DL, Redmond CK, Kavannah M, Cronin WM, Vogel V, Ribidoux A, Dimitrov N, Atkins J, Daly M, Wieland S, Tanchiu E, Ford L, Wolmark N. 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.
- Lesko SM, Mitchell AA. An assessment of the safety of pediatric ibuprofen: A practitioner-based

- randomized clinical trial. *J Am Med Assoc* 1995; **273**: 929–33.
21. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; **3**: 409–20.
 22. Hasford J. Drug risk assessment: a case for large trials with lean protocols. *Pharmacoepidemiol Drug Safety* 1994; **3**: 321–7.
 23. Gray R, Clarke M, Collins R, Peto R. Making randomized trials larger: a simple solution? *Eur J Surg Oncol* 1995; **2**: 137–9.
 24. Buring JE, Hennekens CH. The women's health study: summary of the study design. *J Myocardial Ischemia* 1992; **4**: 27–9.
 25. Hennekens CH, Buring JE. Methodologic considerations in the design and conduct of randomized trials: the U.S. Physicians' Health Study. *Controlled Clin Trials* 1989; **10**: 142S–50S.
 26. O'Brien PC. Data and safety monitoring. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics*. Chichester: Wiley, 1998; 1058–66.
 27. DeMets DL. Data and safety monitoring boards. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics*. Chichester: Wiley, 1998; 1066–71.

34

The Use of Pharmacoepidemiology to Study Beneficial Drug Effects

BRIAN L. STROM

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

KENNETH L. MELMON

Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

INTRODUCTION

In order to be approved for marketing in the United States, drugs must be proven to be safe and effective using “adequate and well-controlled investigations.” Earlier chapters in this book have shown that this premarketing information often is insufficient to provide some of the information about drug toxicity which is clinically most important. The same applies to information about drug efficacy.

In this chapter we will begin by clarifying the different definitions of various types of beneficial drug effect. Then we will discuss the need for *postmarketing studies of drug effectiveness*. Next, we will present the unique methodologic problems raised by studies of beneficial drug effects, as well as potential solutions to these problems. Finally we will evaluate the frequency with which these proposed solutions might be successful. Specific examples of approaches to the study of efficacy also will be presented.

DEFINITIONS

There are at least four different types of measurable drug effect of interest to a prescriber. *Unanticipated harmful effects* are the unwanted effects of drugs that could not have been predicted on the basis of their preclinical pharmacologic profile or the results of premarketing clinical studies. These effects are most often Type B adverse reactions, as defined in Chapter 1. For example, chloramphenicol was not known to cause aplastic anemia at the time it was marketed,¹ nor was the skeletal muscle pain associated with use of HMG-CoA reductase inhibitors. A major research challenge is to discover medically important unanticipated harmful effects as soon as possible after drug marketing. Quantitation of the incidence of these effects is medically useful as well.

Anticipated harmful effects are unwanted effects of drugs that could have been predicted on the basis of preclinical and premarketing studies. They can be either Type A reactions or Type B reactions

(see Chapter 1). One example is the syncope that sometimes occurs after patients take their first dose of prazosin.² Although this effect was known to occur at the time of marketing, a major question remaining to be answered was how often the event occurred. The dominant research challenge that this type of drug effect presents is establishing its incidence.

Unanticipated beneficial effects are desirable effects of drugs that were not anticipated at the time of drug marketing. Although these effects may be medically useful, they are nevertheless side-effects, if they are not the purpose for which the drug was given. An example of an unanticipated beneficial effect is aspirin's ability to decrease the probability of a subsequent myocardial infarction in patients who were given the drug for its analgesic or anti-inflammatory action.³ Only recently 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. 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* is a study of 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* is a study of 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* is a study of 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 35.

Note that the outcome variable for any of these studies can be of multiple different types. They can be clinical outcomes (diseased/undiseased), or so-called "outcomes research," as defined by health services researchers (see Chapter 39 for a discussion of the validity issues involved in measuring such outcomes); they can be measures of quality of life (see Chapter 36), 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 35 and 36); they can be economic outcomes (see Chapter 35); etc. Regardless, the same methodologic issues apply to each.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY 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 34.1).⁴ 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. Compliance during these studies is assured, and the patients included are similar to each other in age and sex, do not have other diseases, and are not taking other drugs. Such restrictions maximize the ability of premarketing studies to demonstrate

Table 34.1. Clinically important information about intended beneficial effects of drugs^a

-
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, and
 - (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.
-

^a Modified from reference 4.

a drug's efficacy, if the drug actually is efficacious. Additional information may then be needed on whether, in the world of daily medical practice, the drug actually achieves the same beneficial effects and whether the drug can and does have other beneficial effects. In addition, at the time of marketing there may be little data on a drug's efficacy relative to other medical or surgical alternatives available for the same indication. Finally, a number of factors that are encountered in the practice of medicine can modify a drug's ability to achieve its beneficial effects. Included are variations in the drug regimen, characteristics of the indication for the drug, and characteristics of the patient receiving the drug, including demographic factors, nutritional status, the presence of concomitant illnesses, the ingestion of drugs, and so on. Many, if not most, of these factors that can influence the effects of drugs are not fully explored prior to marketing.

In order to quantitate the need for postmarketing studies of the beneficial effects of drugs, a comparison was made of the 100 most common drug uses of 1978 (drug–indication pairs) to the

information available to the FDA at the time of its regulatory decisions about the marketing and labeling of the drugs involved in these uses.⁴ The comparison restricted itself to drugs approved after 1962, when the Kefauver–Harris Amendments introduced a requirement for the submission of data about drug efficacy prior to approval of the drug for marketing.

Of the 100 common drug uses, 31 had not been approved by the FDA at the time of initial marketing, and 18 still had not been approved at the time of the comparison. Eight of the 18 unapproved uses were probably medically and therapeutically inappropriate. For example, the use of antibiotics is not justified for the treatment of viral infections, but such use was common. Other unapproved drug–indication pairs could well have been quite appropriate, but the regulatory process does not need to and did not reflect the current medical practice.

Of the 100 common drug uses, eight were based on the assumption that a drug had a particular long-term effect, but only an intermediate effect had been studied prior to marketing. For example, antihypertensive drugs are used for their presumed ability to prevent long-term cardiovascular complications, but are approved for marketing on the basis of their ability to lower blood pressure. Five of the 100 common drug uses may have been for either the intermediate effect or the long-term effect of the drugs, but only the intermediate effect was studied prior to marketing. For example, hypoglycemic agents may be used to control the symptoms of diabetes or to prevent the vascular complications of diabetes, but only the former were studied before drug marketing.

Drugs other than those in the list of 100 common uses were sometimes prescribed as treatment for each of the 52 indications included in those 100 uses. Yet, eight of the uses involved drugs whose effects relative to alternative drugs had not been studied prior to marketing.

The 100 common drug uses also included a number of examples of clinical factors that are able to modify the effects of the drug, but these were not discovered until after drug marketing. Some are listed in Table 34.2. In addition,

Table 34.2. Examples of factors determining drug efficacy that were demonstrated after marketing, selected from the 100 most common drug uses of 1978^a

Factors	Drug	Indication	Comments	Reference
Regimen				
Dose per unit time	Ibuprofen	Rheumatoid arthritis, osteoarthritis	Daily dosage initially approved proved to be suboptimal	5
Distribution of dose over time	Furosemide	Congestive heart failure	Efficacy improved by more frequent, smaller doses	6
Duration	Clonidine	Hypertension	Tolerance develops in the absence of a diuretic	7
	Hypoglycemics (e.g., acetohexamide and tolazamide)	Diabetes mellitus	Tolerance develops in many patients	8
Indication				
Severity	Metaproterenol	Asthma	Patients with severe illness do not have a response without additional, supplementary therapy	9
Subcategories	Desipramine	Depression	May vary with endogenous versus exogenous depression	10
Changes over time	Ampicillin	Otitis media	No longer the drug of choice in some geographic areas due to bacterial resistance	11, 12
Patient				
Age	Diazepam	Anxiety	A given regimen is more effective in the aged than in the young 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	13 14
Other illness	Gentamicin	Infection	Lower doses required in renal failure	15
Other				
Drugs	Lithium	Manic-depressive illness	Clearance impaired by diuretics, e.g., furosemide	16
	Acetohexamide	Diabetes mellitus	Many drugs interfere, by causing hyperglycemia (e.g., diuretics), displacing drug from binding sites (e.g., nonsteroidal anti-inflammatory drugs), etc.	17
Diet	Diuretics (e.g., metolazone, furosemide)	Hypertension	A decrease in sodium intake can improve efficacy	18
	Lithium	Manic-depressive illness	Significant sodium depletion or excess can modify renal excretion	16

^aFrom Reference 4.

additional prescriptions accompanied 62% of the prescriptions studied, and 41% of the prescriptions were for patients who had illnesses other than just the one 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 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 is common.¹⁹⁻²¹

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 should prevent the release of the drug to the marketplace. The data needed for clinical decisions are frequently and understandably different from those needed for regulatory decisions. Studies performed prior to marketing per force are focused predominantly on meeting appropriate regulatory requirements, and only secondarily on providing a basis for optimal therapeutic decisions. The physician also should keep in mind that the FDA is not allowed to regulate physicians but, rather, pharmaceutical manufacturers. This regulation is not aimed at telling a physician precisely how an agent should be used. In addition, the FDA does not initiate its own studies of drug effects, but generally evaluates those submitted to it by manufacturers. Finally, there are reasonable logistical limitations on what can be expected prior to marketing, without undue cost in time and resources, as well as delaying the availability of a chemical entity with a proven potential for efficacy. Thus, it seems that more studies of beneficial drug effects are needed, perhaps as a routine part of postmarketing drug surveillance.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY 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 of the other two and, thereby, can create an apparent association or mask a real one. Studies of intended drug effects present a special methodologic problem of confounding by the indication for therapy.^{22,23} In this case the risk factor under study is the drug and the outcome variable under study is the clinical condition which 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 β -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 β -blocker as part of their usual post-myocardial infarction medical care to patients who were not, measuring the incidence of myocardial infarction in both groups. However, patients with angina, arrhythmias, and hypertension, all indications for β -blocker therapy, all 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 nonsteroidal 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.²⁴ Nevertheless, sometimes the problem of confounding by indication can emerge even in studies of unexpected drug effects (beneficial or harmful). For instance, in a study of hypersensitivity reactions associated with the use of nonsteroidal anti-inflammatory drugs, the increased risk of hypersensitivity reactions evident in patients taking nonsteroidal anti-inflammatory drugs was higher in those using the drugs for acute pain than in those using the drugs for osteoarthritis and other chronic conditions. This probably was because of the intermittent ingestion of the drug by those receiving it for acute pain.²⁵

Although confounding by the indication is a relatively uncommon 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.²⁶ 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 34.3).²² 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 given penicillin to treat pneumococcal pneumonia.

However, in applying this approach, the course of a patient's disease must be sufficiently predictable that one can differentiate a true drug effect from spontaneous improvement. In particular, one must be able to exclude *regression to the mean* as the mechanism of the observed change: individuals selected to participate in a study based upon the severity of their disease spontaneously and usually will tend to improve. One example would be a patient with recurrent headaches. The patient would most likely seek medical attention when the headaches are most severe or most frequent. A spontaneous return to the baseline pattern of headaches generally could be expected. However, if the patient were treated in the interim, then the treating physician likely would view the return to normality as evidence of successful therapy, no matter what treatment was used or whether it contributed anything to the recovery.

Second, some questions about beneficial drug effects can be answered using formal nonexperimental studies, because there is no confounding by the indication. If the decision about whether to treat is not based on a formal indication, but on some other factor that may not be related to the outcome variable under study, such as 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, the traditional epidemiologic techniques of exclusion, matching, stratification, and mathematical modeling can be applied. The indication clearly can be

Table 34.3. Classification of research questions according to their problems of confounding by the indication for therapy^a

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 PPD
4. Confounding by the indication exists and is not controllable	Ampicillin used to treat urinary tract infection

^aFrom reference 22.

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. Recently, researchers have begun to use Propensity Scores towards this end.^{27,28} This is an approach that uses mathematical modeling to predict *exposure*, rather than the traditional approach of predicting *outcome*. This is, essentially, a direct measure of indication. One can then use the propensity score to create categories of probability of exposure, and control for those categories in the analysis. While this approach has many attractive features, especially as a direct way to control for confounding by indication, it is important to point out that it is still dependent on identifying and measuring those variables that are the true predictors of therapeutic choice.

When questions of intended drug effects do not fall into any of the preceding categories, *confounding by the indication cannot be controlled*. Non-experimental study designs cannot then be used, or they can only be used to demonstrate qualitatively some degree of beneficial effect. Specifically, if confounding by the indication is such that treated patients would have a worse clinical outcome than untreated patients, yet the outcome observed in treated patients is better than that observed in untreated patients, some degree of confidence that the drug has a beneficial effect can be built. As an example, patients treated with corticosteroids for status asthmaticus would be expected to be sicker than those not so treated. If patients receiving corticosteroids stop wheezing sooner than those not receiving corticosteroids, corticosteroids would indeed seem to have a beneficial effect. However, if the patients receiving corticosteroids do not stop wheezing sooner than those not receiving corticosteroids, the results of the study are uninterpretable. It is possible that the corticosteroids in fact have no beneficial effect. However, it is also possible that a beneficial effect was present but was being masked by the difference in severity between the two treatment groups.

The qualitative approach illustrated above must be used with caution. First, the effect of the confounding by indication must be opposite in direction to the expected effect of the drug. Second, the effect of the confounding by indication must be absolutely predictable in its direction. Third, the effect of the confounding by indication must be sufficiently large so as to exclude regression to the mean as an explanation for the results. Even if all of these conditions are met, the results must be interpreted only qualitatively, not quantitatively.

Examples of each of these situations are presented in Table 34.3 and discussed further in reference 22.

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 design, if any, could be used to evaluate drug effectiveness.²⁹ Of these 100 "drugs," seven were used in contact lenses and were excluded. The remaining 93 drugs were examined for all potential indications and clinical outcomes that could be used to evaluate intended drug effects. Ultimately we assessed 131 drug uses, that is 131 drug-indication pairs. Each drug use was categorized as to whether a study evaluating the effectiveness of that drug for that indication would present the problem of confounding by the indication and, if so, whether one of the approaches described above would be adequate to address it. Eighty-nine (67.9%) of the drug uses could have been evaluated using simple clinical observations, without formal comparative research. A very few of these drugs were, in fact, approved by FDA on the basis of such studies, 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 seven of the 42 (5.3% of the total), this confounding was not an obstacle to valid nonexperimental research. Most often the validity of the approach rested on the observation that any given physician usually used the drug to treat either all or none of his 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 specifics 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.³⁰ Biochemical studies had documented that the menopause resulted in a negative calcium and phosphorus balance, and that the balance returned towards normal with the ingestion of exogenous estrogens.³¹ Studies of bone density documented that exogenous estrogens prevented the loss of bone density that was associated with the menopause,³² for as long as the estrogens were continued.³³ 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.^{34, 35}

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.³⁶⁻⁵³ 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.³⁰ 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 34.3). 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 34.3. 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.⁴⁷

In fact, the three studies that closely examined the comparability of the study groups were able to document that they were not comparable.^{36, 47, 49} 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.³⁶ 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 followup, 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.⁴⁷

A third study used a case-control design to investigate patients admitted to surgical services.⁴⁹ 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.

Anticoagulants for Prevention of Recurrent Venous Thromboembolism

The use of intravenous anticoagulants reduces the risk of recurrent venous thromboembolism,⁵⁴ and the addition of oral anticoagulants to intravenous anticoagulants probably reduces the risk even further.⁵⁵ However, how long oral anticoagulant treatment should be continued had not been well studied. Most explicit advice from experts on the optimal duration of anticoagulation therapy was and is based on anecdotal experience.^{56,57} Most of the data available that are used to suggest the appropriate duration of therapy are derived from clinical observations in a single medical center.⁵⁸⁻⁶¹ They represent an accumulating case series. Over time, gradually patients' treatment has been prolonged. Thus, changes in the duration of treatment are intermingled with other changes in medical care over decades. In addition, the studies do not compare patients receiving treatments of different length, but simply observe when most recurrences tend to occur. The investigators have assumed that treatment should be prolonged sufficiently to include that time when recurrences can be expected. Problems with these studies have been detailed.^{56,57}

As with the question of the effect of estrogens on bone fractures from osteoporosis, a randomized clinical trial would be the ideal design to address the question of the optimal duration of anticoagulation after venous thromboembolism, but such a study is impractical. After patients have been anticoagulated in the hospital and followed for a short time as outpatients, the risk of recurrence is sufficiently small that an enormous population would be needed to detect a difference in outcome due to differences in therapy. Until recently, the only randomized clinical trial in the literature that addressed this question compared six weeks of outpatient treatment to six months of treatment. No difference in recurrence rate between two groups of patients was observed.⁶² However, only 186 subjects were included, yielding a total of only seven recurrences. In addition, over half the study subjects had known short-term risk factors for venous thromboembolism. These included pregnancy, use of oral contraceptives, and recent surgery. Patients with these transient underlying risk factors might be expected to be less likely to benefit from longer-term anticoagulant therapy than patients with idiopathic disease.

The question of the optimal duration of anticoagulation was addressed in a cohort study using data from the Northern California Kaiser Permanente Medical Program.⁶³ The study required the use of ten years of data from this population of 1.6 million, or a total of 16 million patient years of experience. There were a total of 3384 individuals identified as being hospitalized for venous thromboembolism. Of these, 2473 suffered from idiopathic venous thromboembolism. Their clinical outcomes were evaluated, according to how long they had been treated with oral anticoagulants. Using those treated with six weeks of therapy or less as a control group, prolongation of therapy beyond that point was found to increase the risk of major bleeding dramatically, but to have no effect on recurrence rates.

The feature of this study that allowed the investigators to overcome the problem of confounding by indication was that physician behavior regarding how long therapy was continued

as essentially random (category 3.a.ii in Table 34.3). The choice of how long to treat became random, because there was no prior information on how long one should treat. In fact, the duration of treatment was relatively uniformly distributed across the years of followup, and the results were no different when one restricted the analysis to those who had their anticoagulation stopped because of hemorrhage, rather than at the option of their physician.

Lidocaine for Prevention of Death from Myocardial Infarction

In another study, the efficacy of lidocaine in preventing death from myocardial infarction was studied using a case-control design.⁶⁴ Among patients admitted to a coronary or intensive care unit for acute myocardial infarction, those who died were compared to an equal number of patients who survived. The controls were matched to the cases for age, gender, race, and date of hospitalization. Overall, lidocaine did not protect against death. Lidocaine was effective only when deaths attributable to ventricular arrhythmia were analyzed separately.

In this careful study, the investigators obviously were well aware of the risk of confounding by indication. They attempted to control for this confounding by using the epidemiologic technique of stratification, that is classifying patients according to their risk of dying from myocardial infarction, in order to control for this inequality of risk as a confounding variable. Thus, they treated the study as a category 3.b question in Table 34.3. Unfortunately, however, it is doubtful that one can accurately and fully measure the basis for physicians' judgments about who they think is at high risk of death from myocardial infarction. Similarly, it is unlikely that each individual's risk of dying from a myocardial infarction can be predicted, especially death by ventricular arrhythmia. Certainly a classification according to just the presence or absence of congestive heart failure, as was used, is overly simplistic. In fact, the rates of death attributed to ventricular arrhythmia were virtually identical in those patients with and without congestive heart failure. Nevertheless the

results do coincide with those of a randomized clinical trial evaluating the efficacy of lidocaine in preventing primary ventricular fibrillation.⁶⁵ However, while the drug prevented the arrhythmia in that randomized clinical trial, it did not alter mortality.

Anticoagulants for Prevention of Death from Myocardial Infarction

Whether anticoagulants can prevent death from myocardial infarction had been addressed using randomized clinical trials.⁶⁶ However, the results had been inconsistent and inconclusive, possibly because of problems of sample size. Thus, this question would appear to be a good candidate for a case-control study. Such a study was done,⁶⁷ with the investigators treating this research question as if it were a category 3.b question in Table 34.3. However, as with the study of the effects of lidocaine on myocardial infarction, it is doubtful whether one can measure and quantitate precisely the risk of dying from a myocardial infarction at the time of the acute episode. This study might have been more convincing if the investigators had identified the patients of practitioners who always used anticoagulants for their patients with myocardial infarctions, and then compared them to a control group of patients of practitioners who never used anticoagulants for their patients with myocardial infarctions. Inasmuch as the choice of therapy in these patients would not have been made on the basis of any perceived difference among the patients in their risk of dying from myocardial infarction, confounding by the indication would not be a problem. Of course, if the investigators had designed the study as we suggest, they then would have had to consider whether the physicians themselves were somehow a predictor of outcome, and whether this was consistently related to their philosophy of using anticoagulants, across multiple physicians. Thus, randomized trials are really needed to provide the answer to this question, and of course in recent years, with the advent of low molecular weight heparin and thrombolytic therapy, many have been forthcoming.

Generic versus Brand Name Drugs

Another potential use of nonexperimental study designs to study the beneficial effects of drugs arose with the passage of the 1984 Waxman–Hatch Act in the US. Generic drugs can now be marketed after simple demonstration of bioequivalence i.e., equivalent bioavailability, in 18 to 24 normal adults.⁶⁸ However, it is not clear whether bioequivalence assures clinical equivalence, that is equivalent efficacy and toxicity.⁶⁹ *Clinical inequivalence is more likely to be evident as a difference in beneficial effects than as a difference in adverse effects.* In developing a drug, dosages are sought which optimize drug efficacy. Toxicity, other than idiosyncratic or allergic reactions, usually occurs at higher doses and concentrations than efficacy. Modest variations in the plasma concentration of the active drug, created by receiving the same dose in different preparations, are most likely, therefore, to be a problem for drug efficacy than for drug toxicity. Variations in plasma concentration are even more likely to be a problem for drug effectiveness and cost-effectiveness. Even a simple change in the physical appearance of the drug could conceivably lead to a decrease in compliance and, thereby, effectiveness.

Studies designed to evaluate differences in efficacy among different preparations of the same drug require enormous sample sizes, as one would be searching for relatively small differences. However, such sample sizes can be achieved relatively easily and efficiently as part of nonexperimental pharmacoepidemiology studies. Thus, the suggestion has been made that studies of clinical equivalence could possibly be carried out as postmarketing surveillance studies.⁶⁹ Confounding by the indication is unlikely to be a problem because, as far as the physician is concerned, he or she is dealing with different products of the same drug, products which are theoretically interchangeable. The choice among the alternative therapies is not being made by the prescriber on the basis of patient characteristics, but by the pharmacist on the basis of product availability—category 3.a.ii in Table 34.3.

A few pharmacoepidemiology studies on the relative effectiveness of different preparations used for the same purpose have been performed by Strom, using the COMPASS[®] database. These studies compared patients who began on a brand name product and switched to a generic product when it became available to patients who remained on the brand name product. The drugs studied were thioridazine, chlorpropamide, and slow absorption theophylline. These studies naturally raise concerns about the ability to identify the actual product dispensed. Very few of the pharmacoepidemiology approaches described in Part III of the book are able to identify the specific product dispensed. Often the approach does not even distinguish whether it is a brand name product or a generic product that is being used. Even when the distinction is made, for example most Medicaid datasets use the National Drug Code to identify specifically the drug, the manufacturer, the dosage form, and the dose, one is inevitably left with questions about whether a brand name is being billed for, while a generic drug is dispensed. In addition, such studies raise concerns about how to define the clinical outcome variable. For example, how is drug efficacy reflected in a claims database? The studies described above used outcomes like number of physician visits, number of hospitalizations, and use of adjunctive therapy to obtain an estimate of drug efficacy.

Using these outcomes, the investigators first analyzed the baseline data, comparing the experience, prior to switching, of those who ultimately switched to generic products to the experience of those who did not later switch to a generic product. In each of the three studies, the future switchers were different from the future non-switchers, prior to the switch. Thus, it appears that patients who were to be switched to generic products were different than patients who stayed on the brand name products: confounding by indication was indeed operating. Because of this, no analyses of efficacy after the switch were performed. Parenthetically, because of this, and questions about the uncertain interpretability of the clinical outcomes, it was elected not to publish the results of these papers.

Cost-Effectiveness Studies

An important new 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 35.

Vaccines

In the last 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,^{70,71} rubella vaccine,^{72,73} measles vaccine,^{74–77} *Haemophilus influenzae* type b polysaccharide vaccine,^{78–86} oral poliovirus vaccine,^{87,88} meningococcus vaccine,⁸⁹ Japanese encephalitis vaccine,^{90,91} and BCG vaccine in protecting against tuberculosis^{92–99} and leprosy.^{100,101} Cohort studies have been used to explore the efficacy of *Haemophilus influenzae* type b polysaccharide vaccine,⁷⁹ measles vaccine,⁸⁰ and pertussis vaccine.^{102,103}

Again studies like these should ideally be conducted as randomized clinical trials. However, the relative infrequency of the diseases the above

vaccines are designed to prevent, particularly in populations which are partly vaccinated, make use of this design difficult, although not impossible. In fact, in one situation, a new Chinese manufactured Japanese encephalitis vaccine was studied for efficacy using a case-control design,⁹⁰ despite a study of its safety conducted by the same authors using a randomized clinical trial design.⁹¹ In considering the applicability of nonexperimental study designs, the relatively indiscriminate use of such vaccines places the study in category 2 of Table 34.3. 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. We refer the interested reader to some methodologic papers on the subtleties of designing nonexperimental studies of vaccine efficacy.^{104–110}

Cancer Screening

Another 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 those 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 cancer¹¹¹⁻¹²² and mammography and self-examination for breast cancer.¹²³⁻¹³⁶ Other studies investigated screening measures for lung cancer^{122, 137, 138} and gastric cancer.¹³⁹ All of these were case-control studies. Again, they raise similar methodologic considerations of confounding by indication. Specifically, why do some women choose to have the screening procedure and others do not? One randomized clinical trial documented that women who attended screening sessions were at higher risk of developing breast cancer than women who were offered screening but did not attend.¹⁴⁰ 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.¹⁴¹⁻¹⁵⁶

Other Examples

Other analogous work using case-control study designs has explored the effectiveness of bicycle safety helmets in preventing face injuries,¹⁵⁷ antibiotic prophylaxis in preventing post-dental infective endocarditis,¹⁵⁸ β -blockers in preventing mortality in patients with acute myocardial infarction,¹⁵⁹ β -blockers and incident coronary artery events,¹⁶⁰ etc.

THE FUTURE

Clinicians have long recognized the value of clinical observations and nonexperimental research. Much of our current knowledge about the usefulness of medical interventions is based on information that is nonexperimental. 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."²⁶ Sackett *et al.* argue "... to keep up with the clinical literature ... discard at once all articles on therapy that are not randomized trials."¹⁶¹ In light of the analysis presented above, this posture seems too simplistic and far-reaching. If overbearing, it results in clinically necessary and potentially available information being uncollected and unused. The proper balance in attitude about the value of these approaches probably lies somewhere between the two extremes. To quote Sir Austin Bradford Hill, one of the developers of the randomized trial: "Any belief that the controlled trial is the only way (to study therapeutic efficacy) would mean not only that the pendulum had swung too far but that it had come right off its hook."¹⁶² 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 felt to be unnecessary, unethical, infeasible, or too costly relative to the expected benefits, there frequently is a good alternative.

REFERENCES

1. Wallerstein RO, Condit PK, Kasper CK *et al.* Statewide study of chloramphenicol therapy and fatal aplastic anemia. *J Am Med Assoc* 1969; **208**: 2045-50.
2. Graham RM, Thornell IR, Gain JM, *et al.* Prazosin: the first dose phenomenon. *Br Med J* 1976; **2**: 1293-4.

3. Boston Collaborative Drug Surveillance Group. Regular aspirin intake and acute myocardial infarction. *Br Med J* 1974; **1**: 440–3.
4. Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: Why? *Am J Med* 1985; **78**: 475–80.
5. Anonymous. Ibuprofen (Motrin)—a new drug for arthritis. *Med Lett Drugs Ther* 1974; **16**: 109–10.
6. Wilson TW, Falk KJ, Labelle JL, *et al.* Effect of dosage regimen on natriuretic response to furosemide. *Clin Pharmacol Ther* 1975; **18**: 165–9.
7. Anonymous. Clonidine (Catapres) for hypertension. *Med Lett Drugs Ther* 1975; **17**: 45–6.
8. Shen SW, Bressler R. Clinical pharmacology of oral antidiabetic agents. *New Engl J Med* 1977; **296**: 787–93.
9. Webb-Johnson DC, Andrews JL. Bronchodilator therapy. *New Engl J Med* 1977; **297**: 758–64.
10. Hollister LE. Tricyclic antidepressants. *New Engl J Med* 1978; **299**: 1106–9.
11. Syriopoulou V, Scheifele D, Howie V, *et al.* Incidence of ampicillin-resistant *Hemophilus influenzae* in otitis media. *J Pediatr* 1976; **89**: 839–41.
12. Schwartz R, Rodriquez W, Khan W, *et al.* The increasing incidence of ampicillin-resistant *Hemophilus influenzae*. *J Am Med Assoc* 1978; **239**: 320–3.
13. Vestal RE. Drug use in the elderly: a review of problems and special considerations. *Drugs* 1978; **16**: 358–82.
14. Morselli PL, Principi N, Tognoni G, *et al.* Diazepam elimination in premature and full term infants, and children. *J Perinat Med* 1973; **1**: 133–41.
15. Chan RA, Bennes EJ, Hoepflich PD. Gentamicin therapy in renal failure: a nomogram for dosage. *Ann Intern Med* 1972; **76**: 773–8.
16. Hansten PD. *Drug interactions*, 4th edn Philadelphia, PA: Lea and Febiger, 1979.
17. Physicians' desk reference, 33rd edn Oradell, NJ: Medical Economics, 1979.
18. Parijs J, Joossens JV, Van der Linden L, *et al.* Moderate sodium restriction and diuretics in the treatment of hypertension. *Am Heart J* 1973; **85**: 22–34.
19. Doering PL, Stewart RB. The extent and character of drug consumption during pregnancy. *J Am Med Assoc* 1978; **239**: 843–6.
20. Boethius G. Recording of drug prescriptions in the county of Jämtland, Sweden. II. Drug exposure of pregnant women in relation to course and outcome of pregnancy. *Eur J Clin Pharmacol* 1977; **12**: 37–43.
21. Degenhardt KH, Kerken H, Knörr K. Drug usage and fetal development: preliminary evaluations of a prospective investigation. *Adv Exp Med Biol* 1972; **27**: 467–79.
22. Strom BL, Miettinen OS, Melmon KL. Postmarketing studies of drug efficacy: when must they be randomized? *Clin Pharmacol Ther* 1983; **34**: 1–7.
23. Psaty BM, Koepsell TD, Siscovick D, *et al.* An approach to several problems in using large databases for population-based case-control studies of the therapeutic efficacy and safety of anti-hypertensive medicines. *Stat Med* 1991; **10**: 653–62.
24. Carson JL, Strom BL, Soper KA, *et al.* The association of nonsteroidal anti-inflammatory drugs with upper gastrointestinal tract bleeding. *Arch Intern Med* 1987; **147**: 85–8.
25. Strom BL, Carson JL, Morse ML, *et al.* The effect of indication on hypersensitivity reactions associated with zomepirac sodium and other nonsteroidal antiinflammatory drugs. *Arthritis Rheum* 1987; **30**: 1142–8.
26. Juhl E, Christensen E, Tygstrup N. The epidemiology of the gastrointestinal randomized clinical trial. *New Engl J Med* 1977; **296**: 20–2.
27. Rosenbaum PR. *Springer Series in Statistics: Observational Studies*. New York: Springer, 1995; 200–24.
28. 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.
29. Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? *Am J Med* 1984; **77**: 703–8.
30. Strom BL. Are estrogens effective in preventing fractures from post-menopausal osteoporosis? In: Melmon KL, ed, *Drug Therapeutics: Concepts for Physicians*. New York, Elsevier-North-Holland, 1982, 67–80.
31. Lagrelius A. Treatment with oral estrone sulphate in the female climacteric. III. Effects on bone density and on certain biochemical parameters. *Acta Obstet Gynecol Scand* 1981; **60**: 481–8.
32. Lindsay R, Hart DM, Aitken JM, *et al.* Long-term prevention of postmenopausal osteoporosis by oestrogen. *Lancet* 1976; **1**: 1038–41.
33. Lindsay R, Hart DM, MacLean A, *et al.* Bone response to termination of oestrogen treatment. *Lancet* 1978; **1**: 1325–7.
34. Ziel HK, Finkle WD. Increased risk of endometrial cancer among users of conjugated estrogens. *New Engl J Med* 1975; **293**: 1167–70.
35. Antunes CMF, Stolley PD, Rosenshein NB, *et al.* Endometrial cancer and estrogen use. *New Engl J Med* 1979; **300**: 9–13.
36. Hutchinson TA, Polansky SM, Feinstein AR. Postmenopausal oestrogens protect against fractures of hip and distal radius. *Lancet* 1979; **2**: 705–9.
37. Weiss NS, Ure CL, Ballard JH, *et al.* Decreased risk of fractures of the hip and lower forearm with

- postmenopausal use of estrogen. *New Engl J Med* 1980; **303**: 1195–8.
38. Johnson RE, Specht EE. The risk of hip fracture in postmenopausal females with and without estrogen drug exposure. *Am J Public Health* 1981; **71**: 138–44.
 39. Wallach S, Henneman PH. Prolonged estrogen therapy in post-menopausal women. *J Am Med Assoc* 1959; **171**: 1637–42.
 40. Gordon GS, Picchi J, Roof BS. Antifracture efficacy of long-term estrogens for osteoporosis. *Trans Assoc Am Phys* 1973; **86**: 326–32.
 41. Burch JC, Byrd BF, Vaughn WK. The effects of long-term estrogen on hysterectomized women. *Am J Obstet Gynecol* 1974; **118**: 778–82.
 42. Burch JC, Byrd BF, Vaughn WK. The effects of long-term estrogen administration to women following hysterectomy. *Front Horm Res* 1975; **3**: 208–14.
 43. Nachtigall LE, Nachtigall RH, Nachtigall RD, *et al*. Estrogen replacement therapy I. A 10-year prospective study in the relationship to osteoporosis. *Obstet Gynecol* 1979; **53**: 277–81.
 44. Nachtigall LE, Nachtigall RH, Nachtigall RD, *et al*. Estrogen replacement therapy II. A prospective study in the relationship to carcinoma and cardiovascular and metabolic problems. *Obstet Gynecol* 1979; **54**: 74–9.
 45. Nordin BEC, Horsman A, Crilly RG, *et al*. Treatment of spinal osteoporosis in postmenopausal women. *Br Med J* 1980; **1**: 451–4.
 46. Paganini-Hill A, Ross RK, Gerkins VR, *et al*. Menopausal estrogen therapy and hip fractures. *Ann Intern Med* 1981; **95**: 28–31.
 47. Hammond CB, Jelovsek FR, Lee KL, *et al*. Effects of long-term estrogen replacement therapy I. Metabolic effects. *Am J Obstet Gynecol* 1979; **133**: 525–36.
 48. Riggs BL, Seeman E, Hodgson SF, *et al*. Effect of the fluoride/calcium regimen on vertebral fracture occurrence in postmenopausal osteoporosis. *New Engl J Med* 1982; **306**: 446–50.
 49. Krieger N, Kelsey JL, Holford TR, *et al*. An epidemiologic study of hip fracture in postmenopausal women. *Am J Epidemiol* 1982; **116**: 141–8.
 50. Hooyman JR, Melton LJ, Nelson AM, *et al*. Fractures after rheumatoid arthritis. *Arthritis Rheum* 1984; **27**: 1353–61.
 51. Ettinger B, Genant HK, Cann CE. Long-term estrogen replacement therapy prevents bone loss and fractures. *Ann Intern Med* 1985; **102**: 319–24.
 52. Wasnich RD, Ross PD, Heilbrun LK, *et al*. Differential effects of thiazide and estrogen upon bone mineral content and fracture prevalence. *Obstet Gynecol* 1986; **67**: 457–62.
 53. Michaelsson K, Baron JA, Farahmand BY, Johnell O, Magnusson C, Persson PG, Persson I, Ljunghall S. Hormone replacement therapy and risk of hip fracture: population based case–control study. The Swedish Hip Fracture Study Group. *Br Med J* 1998; **316**: 1858–63.
 54. Barritt DW, Jordan SC. Anticoagulant drugs in the treatment of pulmonary embolism. *Lancet* 1960; **1**: 1309–12.
 55. Coon WW, Willis PW. Recurrence of venous thromboembolism. *Surgery* 1973; **73**: 823–7.
 56. Acheson L, Speizer FE, Tager I. Venous thromboembolism [letter]. *New Engl J Med* 1979; **293**: 879.
 57. Salzman EW. Venous thrombosis. Duration of anticoagulant therapy [letter]. *New Engl J Med* 1975; **293**: 879.
 58. Coon WW, Mackenzie JW, Hodgson PE. A critical evaluation of anticoagulant therapy in peripheral venous thrombosis and pulmonary embolism. *Surg Gynecol Obstet* 1958; **106**: 129–36.
 59. Coon WW, Willis PW, Symons MJ. Assessment of anticoagulant treatment of venous thromboembolism. *Ann Surg* 1969; **170**: 559–68.
 60. Coon WW, Willis PW. Thromboembolic complications during anticoagulant therapy. *Arch Surg* 1972; **105**: 209–12.
 61. Coon WW, Willis PW. Recurrence of venous thromboembolism. *Surgery* 1973; **73**: 823.
 62. O'Sullivan EF. Duration of anticoagulant therapy in venous thrombo-embolism. *Med J Aust* 1972; **2**: 1104–7.
 63. Petitti DB, Strom BL, Melmon KL. Duration of warfarin anticoagulant therapy and the probabilities of recurrent thromboembolism and hemorrhage. *Am J Med* 1986; **81**: 255–9.
 64. Horwitz RI, Feinstein AR. Improved observational method for studying therapeutic efficacy. Suggestive evidence that lidocaine prophylaxis prevents death in acute myocardial infarction. *J Am Med Assoc* 1981; **246**: 2455–9.
 65. Lie KI, Wellens HJ, van Capelle FJ, *et al*. Lidocaine in the prevention of primary ventricular fibrillation. *New Engl J Med* 1974; **291**: 1324–6.
 66. Hirsh J. Effectiveness of anticoagulants. *Semin Thromb Hemost* 1986; **12**: 21–37.
 67. Horwitz RI, Feinstein AR. The application of therapeutic-trial principles to improve the design of epidemiologic research: a case–control study suggesting that anticoagulants reduce mortality in patients with myocardial infarction. *J Chron Dis* 1981; **34**: 575–83.
 68. Mattison N. Pharmaceutical innovation and generic drug competition in the USA: effects of the Drug Price Competition and Patent Term Restoration Act of 1984. *Pharmaceut Med* 1986; **1**: 177–85.
 69. Strom BL. Generic drug substitution–revisited. *New Engl J Med* 1987; **316**: 1456–62.

70. Shapiro ED, Clemens JD. A controlled evaluation of the protective efficacy of pneumococcal vaccine for patients at high risk of serious pneumococcal infections. *Ann Intern Med* 1984; **101**: 325–30.
71. Rheingold AL, Broome CV, Hightower AW, *et al*. Age-specific differences in duration of clinical protection after vaccination with meningococcal polysaccharide A vaccine. *Lancet* 1985; **2**: 114–21.
72. Greaves WL, Orenstein WA, Hinman AR, *et al*. Clinical efficacy of rubella vaccine. *Pediatr Infect Dis* 1983; **2**: 284–6.
73. Strassburg MA, Greenland S, Stephenson TG, *et al*. Clinical effectiveness of rubella vaccine in a college population. *Vaccine* 1985; **3**: 109–12.
74. Killewo J, Makwaya C, Munubhi E, Mpenbeni R. The protective effect of measles vaccine under routine vaccination conditions in Dar Es Salaam, Tanzania: a case-control study. *Int J Epidemiol* 1991; **20**: 508–14.
75. Mahomva AI, Moyo IM, Mbengeranwa LO. Evaluation of a measles vaccine efficacy during a measles outbreak in Mbare, City of Harare Zimbabwe. *Cent Afr J Med* 1997; **43**: 254–6.
76. Mudzamiri WS, Peterson DE, Marufu T, Biellik RJ, L'Hermines M. Measles vaccine efficacy in Masvingo district, Zimbabwe. *Cent Afr J Med* 1996; **42**: 195–7.
77. Wenger JD, Pierce R, Deaver KA, Plikaytis BD, Facklam RR, Broome CV and the Haemophilus Influenzae Vaccine Efficacy Study Group. Efficacy of Haemophilus influenzae type b polysaccharide-diphtheria toxoid conjugate vaccine in US children aged 18–59 months. *Lancet* 1991; **338**: 395–8.
78. Greenberg DP, Vadheim CM, Bordenave N, Ziontz L, Christenson P, Waterman SH, Ward JI. Protective efficacy of Haemophilus influenzae type b polysaccharide and conjugate vaccines in children 18 months of age and older. *J Am Med Assoc* 1991; **265**: 987–92.
79. Kaiser Permanente Pediatric Vaccine Study Group. Efficacy of Haemophilus influenzae type b capsular polysaccharide vaccine. *Pediatr Infect Dis* 1988; **7**: 149–56.
80. Sutcliffe PA, Rea E. Outbreak of measles in a highly vaccinated secondary school population. *Can Med Assoc J* 1996 **155**: 1407–13.
81. Harrison LH, Broome CV, Hightower AW, Hoppe CC, Makintubee S, Sitze SL, Taylor JA, Gaventa S, Wenger JD, Facklam RR, and the Haemophilus Vaccine Efficacy Study Group. A day care-based study of the efficacy of Haemophilus b polysaccharide vaccine. *J Am Med Assoc* 1988; **260**: 1413–8.
82. Shapiro ED, Murphy TV, Wald ER, Brady CA. The protective efficacy of Haemophilus b polysaccharide vaccine. *J Am Med Assoc* 1988; **260**: 1419–22.
83. Osterholm MT, Rambeck JH, White KE, Jacobs JL, Pierson LM, Neaton JD, Hedberg CW, MacDonald KL, Granoff DM. Lack of efficacy of Haemophilus b polysaccharide vaccine in Minnesota. *J Am Med Assoc* 1988; **260**: 1423–8.
84. Black SB, Shinefield HR, Hiatt RA, Fireman BH, and the Kaiser Permanente Pediatric Vaccine Study Group. Efficacy of Haemophilus influenzae type b capsular polysaccharide vaccine. *Pediatr Infect Dis J* 1988; **7**: 149–56.
85. Bower C, Condon R, Payne J, Burton P, Watson C, Wild B. Measuring the impact of conjugate vaccines on invasive Haemophilus influenzae type b infection in Western Australia. *Aust N Z J Public Health* 1998; **22**: 67–72.
86. Muhlemann K, Alexander ER, Weiss NS, Pepe M, Schopfer K. Risk factors for invasive Haemophilus influenzae disease among children 2–16 years of age in the vaccine era, Switzerland 1991–1993. The Swiss H. Influenzae Study Group. *Int J Epidemiol* 1996; **25**: 1280–5.
87. Balraj V, John TJ, Thomas M, Mukundan S. Efficacy of oral poliovirus vaccine in rural communities of North District, India. *Int J Epidemiol* 1990; **19**: 711–4.
88. Deming MS, Jaiteh KO, Otten MW Jr, *et al*. Epidemic poliomyelitis in The Gambia following the control of poliomyelitis as an endemic disease. *Am J Epidemiol* 1992; **135**: 393–408.
89. Rosenstein N, Levine O, Taylor JP, Evans D, Plikaytis BD, Wenger JD, Perkins BA. Efficacy of meningococcal vaccine and barriers to vaccination. *J Am Med Assoc* 1998; **279**: 435–9.
90. Hennessy S, Lui Z, Tsai TF, Strom BL, Wan CM, Liu HL, Wu TX, Yu HJ, Liu QM, Karabatsos N, Bilker WB, Halstead SB. Effectiveness of live-attenuated Japanese encephalitis vaccine (SA14–14–2): a case-control study. *Lancet* 1996; **347**: 1583–6.
91. Liu ZL, Hennessy S, Strom BL, Tsai TF, Wan CM, Tang SC, Xiang CF, Bilker WB, Pan XP, Yao YJ, Xu ZW, Halstead SB. Short-term safety of live-attenuated Japanese Encephalitis vaccine (SA14–14–2): results of a 26,239-subject randomized clinical trial. *J Infect Dis* 1997; **176**: 1366–9.
92. Miceli I, Colaiacovo D, Kantor IN, *et al*. Evaluation of the effectiveness of BCG vaccination using the case-control method. *Dev Biol Stand* 1986; **58**: 293–6.
93. Shapiro C, Cook N, Evans D, *et al*. A case-control study of BCG and childhood tuberculosis in Cali, Colombia. *Int J Epidemiol* 1985; **14**: 441–6.
94. Young TK, Hershfield ES. A case-control study to evaluate the effectiveness of mass BCG vaccination among Canadian Indians. *Am J Public Health* 1986; **76**: 783–6.

95. Houston S, Fanning A, Soskolne CL, Fraser N. The effectiveness of bacillus Calmette-Guérin (BCG) vaccination against tuberculosis. *Am J Epidemiol* 1990; **131**: 340-8.
96. Filho VW, de Castilho EA, Rodrigues LC, Huttly SRA. Effectiveness of BCG vaccination against tuberculous meningitis: a case-control study in Sao Paulo, Brazil. *Bull WHO* 1990; **68**: 69-74.
97. Patel A, Schofield F, Siskind V, *et al.* Case-control evaluation of a school-age BCG vaccination programme in subtropical Australia. *Bull WHO* 1991; **69**: 425-33.
98. Sirinavin S, Chotpitayusunondh T, Suwanjutha S, *et al.* Protective efficacy of neonatal Bacillus Calmette-Guérin vaccination against tuberculosis. *Pediatric Infect Dis J* 1991; **10**: 359-65.
99. Sharma RS, Srivastava DK, Singh AA, *et al.* Epidemiological evaluation of BCG vaccine efficacy in Delhi—1989. *J Commun Dis* 1989; **21**: 200-6.
100. Fine PEM, Ponninghaus JM, Maine N, *et al.* Protective efficacy of BCG against leprosy in Northern Malawi. *Lancet* 1986; **2**: 499-502.
101. Muliyl J, Nelson KE, Diamond EL. Effect of BCG on the risk of leprosy in an endemic area: a case control study. *Int J Lepr* 1991; **59**: 229-36.
102. Isomura S. Clinical studies on efficacy and safety of an acellular pertussis vaccine in Aichi prefecture, Japan. *Dev Biol Stand* 1991; **73**: 37-42.
103. Report from the PHLS Epidemiological Research Laboratory and 21 area health authorities. Efficacy of pertussis vaccination in New England. *Br Med J* 1982; **285**: 357-9.
104. Smith PG. Retrospective assessment of the effectiveness of BCG vaccination against tuberculosis using the case-control method. *Tubercle* 1982; **63**: 23-35.
105. 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.
106. Clemens JD, Shapiro ED. Resolving the pneumococcal vaccine controversy: are there alternatives to clinical trials? *Rev Infect Dis* 1984; **6**: 589-600.
107. Orenstein WA, Bernier RH, Dondero TJ, *et al.* Field evaluation of vaccine efficacy. *Bull WHO* 1985; **63**: 1055-68.
108. Smith PG. Evaluating interventions against tropical diseases. *Int J Epidemiol* 1987; **16**: 159-66.
109. Farrington CP. Quantifying misclassification bias in cohort studies of vaccine efficacy. *Stat Med* 1990; **9**: 1327-37.
110. Joseph KS. A potential bias in the estimation of vaccine efficacy through observational study designs. [letter]. *Int J Epidemiol* 1989; **18**: 729-31.
111. Clarke EA, Anderson TW. Does screening by "Pap" smears help prevent cervical cancer? *Lancet* 1979; **2**: 1-4.
112. Aristizabal N, Cuello C, Correa P, *et al.* The impact of vaginal cytology on cervical cancer risks in Cali, Colombia. *Int J Cancer* 1984; **34**: 5-9.
113. LaVecchia C, Franceschi S, Decarli A, *et al.* "Pap" smear and the risk of cervical neoplasia: quantitative estimates from a case-control study. *Lancet* 1984; **2**: 779-82.
114. MacGregor JE, Moss SM, Parkin DM, *et al.* A case-control study of cervical cancer screening in north east Scotland. *Br Med J* 1985; **290**: 1543-6.
115. van der Graaf Y, Zielhuis GA, Peer PGM *et al.* The effectiveness of cervical screening: a population-based case-control study. *J Clin Epidemiol* 1988; **41**: 21-6.
116. Clarke EA, Hilditch S, Anderson TW. Optimal frequency of screening for cervical cancer: a Toronto case-control study. *IARC Sci Publ* 1986; **76**: 125-31.
117. Raymond L, Obradovic M, Riotton G. Additional results on relative protection of cervical cancer screening according to stage of tumour from the Geneva case-control study. *IARC Sci Publ* 1986; **76**: 107-10.
118. Geirsson G, Kristiansdottir R, Sigurdsson K, *et al.* Cervical cancer screening in Iceland: a case-control study. *IARC Sci Publ* 1986; **76**: 37-41.
119. Berrino F, Gatta G, d'Alto M, *et al.* Efficacy of screening in preventing invasive cervical cancer: a case-control study in Milan, Italy. *IARC Sci Publ* 1986; **76**: 111-23.
120. Celentano DD, Klassen AC, Weisman CS, Rosenhein NB. Cervical cancer screening practices among older women: results from the Maryland cervical cancer case-control study. *J Clin Epidemiol* 1988; **41**: 531-41.
121. Zhang ZF, Parkin DM, Yu SZ, *et al.* Cervical screening attendance and its effectiveness in a rural population in China. *Cancer Detect Prev* 1989; **13**: 337-42.
122. Sobue T, Suzuki T, Fujimoto I, *et al.* Population-based case-control study on cancer screening. *Environ Health Perspect* 1990; **87**: 57-62.
123. Collette JHA, Day NE, Rombach JJ, de Waard F. Evaluation of screening for breast cancer in a non-randomized study (the DOM project) by means of a case-control study. *Lancet* 1984; **1**: 1224-6.
124. Foster RS Jr, Costanza MC. Breast self-examination practices and breast cancer survival. *Cancer* 1984; **53**: 999-1005.
125. Greenwald P, Nasca PC, Lawrence CE, *et al.* Estimated effect of breast self-examination and routine physical examinations on breast cancer mortality. *New Engl J Med* 1978; **299**: 271-3.
126. Smith EM, Francis AM, Polissar L. The effect of breast self-exam practices and physical examina-

- tion on extent of disease at diagnosis. *Prev Med* 1980; **9**: 409–17.
127. Huquely CM, Brown RL. The value of breast self-examination. *Cancer* 1981; **47**: 989–95.
128. Senie RT, Rosen PP, Lesser ML, Kinne DW. Breast self-examination and medical examination related to breast cancer stage. *Am J Public Health* 1981; **71**: 583–90.
129. Feldman JG, Carter AC, Nicastrì AD, et al. Breast self-examination, relationship to stage of breast cancer at diagnosis. *Cancer* 1981; **47**: 2740–5.
130. Dubin N, Pasternack BS. Breast cancer screening data in case-control studies. *Am J Epidemiol* 1984; **120**: 8–16.
131. Dubin N, Friedman DR, Toniolo PG, et al. Breast cancer detection centers and case-control studies of the efficacy of screening. *J Chronic Dis* 1987; **40**: 1041–50.
132. Verbeek ALM, Hendriks JHCL, Holland R, et al. Mammographic screening and breast cancer mortality: age-specific effects in Nijmegen project, 1975–82. *Lancet* 1985; **1**: 865–6.
133. Palli D, Del Turco MR, Buiatti E, Carli S, et al. A case-control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). *Int J Cancer* 1986; **38**: 501–4.
134. Palli D, Del Turco MR, Buiatti E, Ciatto S, et al. Time interval since last test in a breast cancer screening programme: a case-control study in Italy. *J Epidemiol Community Health* 1989; **43**: 241–8.
135. Friedman DR, Dubin N. Case-control evaluation of breast cancer screening efficacy. *Am J Epidemiol* 1991; **133**: 974–84.
136. Verbeek ALM, Holland R, Sturmans F, et al. Reduction of breast cancer mortality through mass screening with modern mammography. *Lancet* 1984; **1**: 1222–4.
137. Ebeling K, Nischan P. Screening for lung cancer—results from a case-control study. *Int J Cancer* 1987; **40**: 141–4.
138. Sobue T, Suzuki T, Naruke T, and the Japanese Lung-Cancer-Screening Research Group. A case-control study for evaluating lung-cancer screening in Japan. *Int J Cancer* 1992; **50**: 230–7.
139. Oshima A, Hirata N, Ubukata T, et al. Evaluation of a mass screening program for stomach cancer with a case-control study design. *Int J Cancer* 1986; **38**: 829–33.
140. Shapiro S, Venet W, Strax P, et al. Selection, follow-up, and analysis in the Health Insurance Plan Study: a randomized trial with breast cancer screening. *J Natl Cancer Inst Monogr* 1985; **67**: 65–74.
141. 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.
142. Morrison AS. Case definition in case-control studies of the efficacy of screening. *Am J Epidemiol* 1982; **115**: 6–8.
143. Weiss NS. Control definition in case-control studies of the efficacy of screening and diagnostic testing. *Am J Epidemiol* 1983; **118**: 457–60.
144. Sasco AJ, Day NE, Walter SD. Case-control studies for the evaluation of screening. *J Chronic Dis* 1986; **39**: 399–405.
145. Frommer DJ. Case-control studies of screening. *J Clin Epidemiol* 1988; **41**: 101.
146. Sasco AJ. Lead time and length bias in case-control studies for the evaluation of screening. *J Clin Epidemiol* 1988; **41**: 103–4.
147. Cole P, Morrison AS. Basic issues in population screening for cancer. *J Natl Cancer Inst* 1980; **64**: 1263–72.
148. Morrison AS. Case definition in case-control studies of the efficacy of screening. *Am J Epidemiol* 1982; **115**: 6–8.
149. Baum M, MacRae KD. Screening for breast cancer. *Lancet* 1984; **2**: 462.
150. Flanders WD, Longini IM Jr. Estimating benefits of screening from observational cohort studies. *Stat Med* 1990; **9**: 969–80.
151. 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.
152. Knox G. Case-control studies of screening procedures. *Public Health* 1991; **105**: 55–61.
153. 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.
154. Weiss NS. Analysis of case-control studies of the efficacy of screening for cancer: how should we deal with tests done in persons with symptoms? *Am J Epidemiol* 1998; **147**: 1099–102.
155. 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.
156. Weiss NS, Lazovich D. Case-control studies of screening efficacy: the use of persons newly diagnosed with cancer who later sustain an unfavorable outcome. *Am J Epidemiol* 1996; **143**: 319–22.
157. Thompson DC, Thompson RS, Rivara FP, Wolf ME. A case-control study of the effectiveness of bicycle safety helmets in preventing facial injury. *Am J Public Health* 1990; **80**: 1471–4.
158. Imperiale TF, Horwitz RI. Does prophylaxis prevent postdental infective endocarditis? A controlled evaluation of protective efficacy. *Am J Med* 1990; **88**: 131–6.
159. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; **89**: 630–8.

160. Psaty BM, Koepsell TD, LoGerfo JP, *et al.* β -blockers and primary prevention of coronary heart disease in patients with high blood pressure. *J Am Med Assoc* 1989; **261**: 2087–94.
161. Sackett DL, Haynes RB, Tugwell PT. *Clinical Epidemiology. A Basic Science for Clinical Medicine*. Boston: Little Brown, 1985.
162. Hill AB. Reflections on the controlled trial. *Ann Rheum Dis* 1966; **25**: 107–13.

35

Pharmacoeconomics: Economic Evaluation of Pharmaceuticals*

KEVIN A. SCHULMAN

Center for Clinical and Genetic Economics, Duke Clinical Research Institute, Duke University Medical Center,
Durham, NC, USA

HENRY GLICK

Leonard Davis Institute of Health Economics, University of Pennsylvania School of Medicine, Philadelphia,
PA, USA

DANIEL POLSKY

Leonard Davis Institute of Health Economics, University of Pennsylvania School of Medicine, Philadelphia,
PA, USA

JOHN M. EISENBERG

Agency for Healthcare Research and Quality, United States Department of Health and Human Services,
Washington, DC, USA

INTRODUCTION

Conventional evaluation of new medical technologies such as pharmaceutical products includes consideration of efficacy, effectiveness, and safety. Other chapters of this book describe in detail how such evaluations are carried out. The methodology is well developed, and federal regulation requires studies of safety and efficacy to be performed prior to drug marketing. Recently, health care researchers from a variety of disciplines have developed

new techniques for the evaluation of the economic effects of clinical care and new medical technologies. Clinicians, pharmacists, economists, epidemiologists, operations researchers, and others have contributed to a new field of “clinical economics,” an evolving discipline dedicated to the study of how different approaches to patient care and treatment influence the resources consumed in clinical medicine.^{1–14}

The growth of clinical economics has proceeded rapidly as health policymakers have faced a

* This chapter does not necessarily represent the views of DHHS or the Federal government.

continuing series of decisions about funding new clinical therapies in an era of increasingly constrained health care resources. Assessments of new therapies include the resources required for the new therapy, the extent of the substitution of the new resources for existing resources, if any, and the health outcomes that result from therapeutic intervention. Thus, clinical economics includes not just an assessment of the cost of a new therapy, but an assessment of its overall economic and clinical effect.

This chapter discusses the need for applying economic concepts to the study of pharmaceuticals, introduces the concepts of clinical economics and the application of these concepts to pharmaceutical research, reviews some of the methodologic issues addressed by investigators studying the economics of pharmaceuticals, and finally offers examples of this type of research.

CLINICAL PROBLEMS TO BE SOLVED BY PHARMACOEPIDEMOLOGY RESEARCH

Recent years have seen increasing 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. Australia is implementing and Canada is considering a set of national guidelines that would mandate the presentation of pharmacoeconomic data at the time of product registration for pharmaceuticals to qualify for reimbursement through the national health insurance systems.^{15, 16} 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 twofold: (i) understanding the potential effect of a therapy (e.g., whether a new antiseptic 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 issues related to the transition from efficacy to efficiency in clinical practice.^{17, 18} 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. 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 due to the regulatory process of drug approval. After a compound is identified and thought to be clinically useful, four distinct sets of evaluations—referred to as phase I through IV studies—are mandated by the US Food and Drug Administration (FDA) and most other equivalent regulatory bodies. Phase I studies represent the first introduction of a new compound into (usually undiseased) humans, principally for the evaluation of safety and dosage. In phase II studies, the drug is introduced into a patient population with the disease of interest, again principally for the evaluation of safety and dosing. Phase III studies are randomized trials evaluating the safety and efficacy of new drugs, compared either with placebo or with a therapy that the new drug might replace (in the US, the appropriate comparator often is the subject of negotiations between the developer of the drug and the FDA). In addition to these three types of study, drugs often are evaluated after they are marketed in what are referred to as phase IV or postmarketing studies. The drug development process allows for timely collection of data that can be used to evaluate the

costs and effects of pharmaceuticals early in their product life, with an opportunity for further data collection and evaluation once the product has been approved and marketed.

Clinical economics is now being integrated throughout the development process, with goals that parallel the clinical development stages. Phase I and II studies are used to develop pilot economic data, such as estimates of the mean and variance estimates for costs, quality of life, and utilities for patients with a specific clinical syndrome. These studies are also used to perform pilot tests of data collection tools, including economic case report forms that prospectively capture resources used by patients who will be entered into the phase III and postmarketing clinical trials. From these data, issues such as sample size and power for pharmacoeconomic studies can be assessed.

One of the fastest growing areas in the economic assessment of new drugs is the incorporation of economic analyses as part of phase III clinical trials. Phase III studies can include economic assessments of new therapies as a primary or secondary endpoint (i.e., an assessment of changes in the use of specific resource categories resulting from treatment, such as changes in length of hospital stay or changes in hospitalization rates).

Lastly, a wide variety of postmarketing economic studies can be performed. These include efficiency trials and postmarketing surveillance studies. In efficiency trials, comparisons between products are made in more realistic settings with less restrictive protocols than those designed for phase III safety and efficacy trials. These postmarketing studies may include assessments of the new therapy compared with “usual care” or compared with specific therapeutic agents. In postmarketing surveillance studies, observational data may be used to evaluate costs, effectiveness, and adverse experiences related to the drug. Again, the economic analysis can serve as a primary or secondary endpoint of the study.

Developing economic data as an endpoint in a clinical trial requires integrating pharmacoeconomics into the clinical development process. While there recently has been an increase in the number of trials that collect economic data, the

challenge remains to ensure that pharmaco-economic endpoints are considered sufficiently early in the clinical development process so that designing the economic protocol does not impede the process of designing the clinical trial. Economic analysis requires the establishment of a set of economic endpoints for study (e.g., direct, productivity, and intangible costs to patients and caregivers, as well as quality of life or preference measures for patients and caregivers), review of the clinical protocol to ensure that there are no economic biases in the design of the clinical trial, and the development of the economic protocol. Ideally, the economic study will be integrated into the clinical protocol and the economic data will be collected as part of a unified case report form for both clinical and economic variables.

In the following sections, we briefly review the research methods of pharmaco-economics, discuss some methodologic issues that have confronted researchers investigating the economics of pharmaceuticals, and review several studies that illustrate the usefulness of pharmaco-economic research.

METHODOLOGIC PROBLEMS TO BE SOLVED BY PHARMACOEPIDEMOLOGY RESEARCH

TECHNIQUES OF CLINICAL ECONOMICS

Economists emphasize that costs are more than just transactions of currency. Cost represents the consumption of a resource that could otherwise be used for another purpose. The value of the resource is that of its next best use, which no longer is possible once the resource has been used. This value is called the resource's "opportunity cost." For example, the time it takes to read this chapter is a cost for the reader, because it is time that cannot be used again; the opportunity to use it for another purpose has been foregone. Good investments are made when the benefits of the investment (e.g., what you learn) are greater than or equal to the value of the opportunities you have forgone (e.g., what you would be doing if you were not reading this chapter).

In addition to the fact that not all costs involve a transaction of money, it is important to remember that, at least from the perspective of society as a whole, not all transactions of money should be considered costs. For example, monetary transactions that do not represent the consumption of resources (e.g., social security payments, disability payments, or other retirement benefits) are not costs by this definition. They simply transfer the right to consume the resources represented by the money from one individual to another.

In considering economic analysis of medical care, there are three dimensions of analysis (represented by the three axes of the cube in Figure 35.1) with which readers should become familiar.¹ Along the *x* axis are three types of economic analysis—cost identification, cost-effectiveness, and cost-benefit. Along the *y* axis are four points of view, or perspectives, that one may take in carrying out an analysis. One may take the point of view of society in assessing the costs and benefits of a new medical therapy. Alternatively, one may take the point of view of the patient, the payer, or the provider. Along the third axis, the *z* axis, are the types of cost and benefit that can be included in economic analysis of medical care. These costs and benefits, which will be defined below, include direct costs and benefits, productivity costs and benefits, and intangible costs and benefits.

Types of Analysis

Cost-benefit Analysis

Cost-benefit analysis of medical care compares the cost of a medical intervention to its benefit. Both costs and benefits are measured in the same (usually monetary) units (e.g., dollars). These measurements are used to determine either the ratio of dollars spent to dollars saved or the net saving (if benefits are greater than costs) or net cost. All else equal, an investment should be undertaken when its benefits exceed its costs.

Costs and benefits are considered to be either positive or negative. A cost can be incurred or avoided and a benefit can be gained or lost. Thus, the cost-benefit ratio is open to manipulation. For

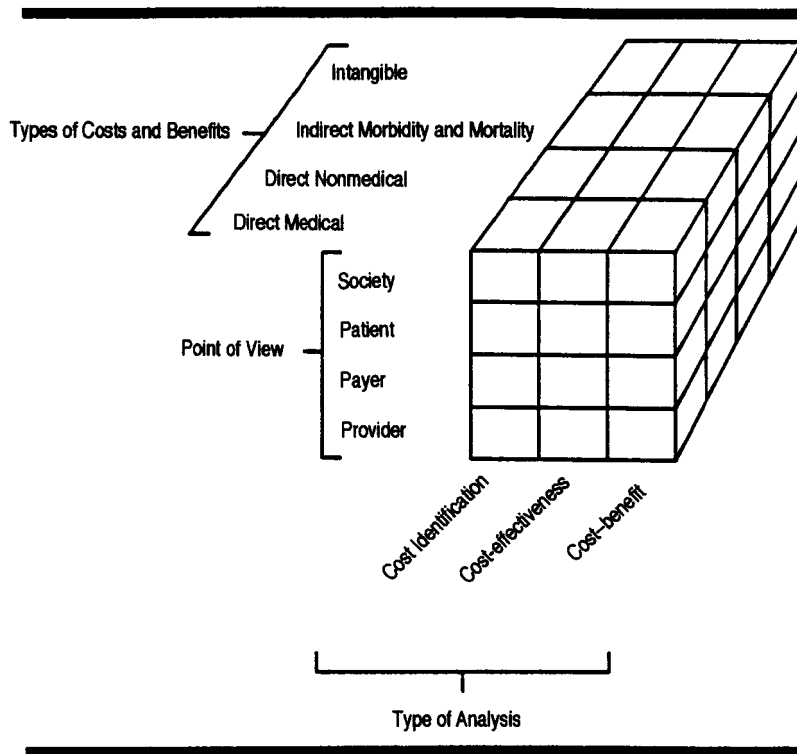


Figure 35.1. The three dimensions of economic evaluation of clinical care (from reference 1).

example, by reclassifying a cost that is incurred as a lost benefit, one can move the term from the numerator, where it would be considered a cost, to the denominator, where it would be considered a negative benefit. The net savings calculation where cost and benefits are summed together cannot be manipulated in this way and is, therefore, generally preferred.

The methods of cost-benefit analysis may be applied to evaluate the total costs and benefits of the interventions that are being compared by analyzing their cost-benefit ratios or their net benefits. Furthermore, the additional or “incremental” cost of an intervention (i.e., the difference in cost between a new therapy and conventional medical care) may be compared with its additional or “incremental” benefit. Incremental analysis is generally preferred to comparisons of totals because it allows the analyst to focus on the differences between any two treatment modalities.

One potential difficulty of cost-benefit analysis is that it requires researchers to express an intervention’s costs and outcomes in the same units. Thus, monetary values must be associated with years of life lost and morbidity due to disease and with years of life gained and morbidity avoided due to intervention. Expressing costs in this way is obviously difficult in health care analyses. Outcomes (treatment benefits) may be difficult to measure in units of currency. Translating disease and treatment outcomes into monetary measures may be more difficult than translating them into clinical outcome measures, such as years of life saved or years of life saved adjusted for quality.

Cost-effectiveness Analysis

Cost-effectiveness analysis provides an approach to the dilemma of assessing the monetary value of health outcomes as part of the evaluation. While

cost generally is still calculated only in terms of dollars spent, effectiveness is determined independently and may be measured only in clinical terms, using any meaningful clinical unit. For example, one might measure clinical outcomes in terms of number of lives saved, complications prevented, or diseases cured. Alternatively, health outcomes can be reported in terms of a change in an intermediate clinical outcome, such as cost per % change in blood cholesterol level. These results generally are reported as a ratio of costs to clinical benefits, with costs measured in monetary terms but with benefits measured in the units of the relevant outcome measure (for example, dollars per year of life saved).

When several outcomes result from a medical intervention (e.g., the prevention of both death and disability), cost-effectiveness analysis may consider these two outcomes together only if a common measure of outcome can be developed. Frequently, analysts combine different categories of clinical outcomes according to their desirability, assigning a weighted utility, or value, to the overall treatment outcome.³ A utility weight is a measure of the patient's preferences for her health state or for the outcome of an intervention. The comparison of costs and utilities sometimes is referred to as cost-utility analysis.

As with cost-benefit analysis, cost-effectiveness analysis can compare a treatment's total costs and total effectiveness, or it can assess only the treatment's incremental costs and incremental effectiveness. In the former, the cost-effectiveness ratio of each intervention is calculated and the two ratios are compared (e.g., the cost per life saved using each intervention). In the latter approach, which assesses incremental costs and benefits, the incremental cost of the innovation is calculated, as is the incremental effectiveness, and the analyst can calculate the additional effect (e.g., lives saved) per additional treatment dollar spent. Programs that cost less and demonstrate improved or equivalent treatment outcomes are said to be dominant and should always be adopted. Programs that cost more and are more effective should be adopted if both their cost-effectiveness and incremental cost-effectiveness ratios fall within an acceptable range and the budget for the program is

acceptable. Programs that cost more and have worse clinical outcomes are said to be dominated and should never be adopted. Programs that cost less and have reduced clinical outcomes may be adopted depending upon the magnitude of the changes in cost and outcome.

As with the translation of clinical outcomes into monetary measures, there also are difficulties associated with combining different outcomes into a common measure in cost-effectiveness analysis. However, it generally is considered more difficult to translate all health benefits into monetary units for the purposes of cost-benefit analysis than to combine clinical outcomes measures. Thus, cost-effectiveness analysis is used more frequently than cost-benefit analysis in the medical care literature.

Cost-identification Analysis

An even less complex approach than cost-benefit or cost-effectiveness analysis would be simply to enumerate the costs involved in medical care and to ignore the outcomes that result from that care. This approach is known as cost-identification analysis. By performing cost-identification analysis, the researcher can determine alternative ways of providing a service. The analysis might be expressed in terms of the cost per unit of service provided. For example, a cost-identification study might measure the cost of a course of antibiotic treatment, but it would not calculate the clinical outcomes (cost-effectiveness analysis) or the value of the outcomes in units of currency (cost-benefit analysis). Cost-identification studies, which include comparisons among different treatments based upon their costs alone, are appropriate only if treatment outcomes or benefits are equivalent for the therapies being evaluated.

Sensitivity Analysis

Most cost-benefit and cost-effectiveness studies require large amounts of data that may vary in reliability, validity, or the effect on the overall results of the study. This is especially the case when models are developed for the economic analysis using secondary data sources, when data collection is performed retrospectively, or when critical data

elements are unmeasured or unknown. Sensitivity analysis is a set of procedures in which the results of a study are recalculated using alternate values for some of the study's variables in order to test the sensitivity of the conclusions to these altered specifications. Such an analysis can yield several important results by demonstrating the independence or dependence of a result on particular assumptions, establishing the minimum or maximum values of a variable that would be required to affect a recommendation to adopt or reject a program, and identifying clinical or economic uncertainties that require additional research. In general, sensitivity analyses are performed on variables that have a significant effect on the study's conclusions but for which values are uncertain.

Types of Costs

Another dimension of economic analysis of clinical practice illustrated by Figure 35.1 is the evaluation of costs of a therapy. Economists consider three types of cost—direct, productivity, and intangible.

Direct Medical Costs

The direct medical costs of care usually are associated with monetary transactions and represent costs that are incurred during the provision of care. Examples of direct medical costs include payments for purchasing a pharmaceutical product, payments for physicians' fees, salaries of allied health professionals, or purchases of diagnostic tests. Because the charge for medical care may not accurately reflect the resources consumed, accounting or statistical techniques may be needed to determine direct costs.^{7, 19–22}

Direct Nonmedical Costs

Monetary transactions undertaken as a result of illness or health care to detect, prevent, or treat disease are not limited to direct medical costs. There is another type of cost that often is overlooked—direct nonmedical costs. These costs are incurred because of illness or the need to seek medical care. They include the cost of transportation to the hospital or physician's office, the cost of

special clothing needed because of the illness, the cost of hotel stays for receiving medical treatment at a distant medical facility, and the cost of special housing (e.g., the cost of modification of a home to accommodate an ill individual). Direct nonmedical costs, which are generally paid out of pocket by patients and their families, are just as much direct medical costs as are expenses that are more usually covered by third party insurance plans.

Direct medical costs can be further classified to help determine the potential effect of a therapy in terms of the ability to change patterns of resource consumption by patients. If these 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. If enough increase in activity occurs, a new building might be needed. Alternatively, if patient care is transferred from an inpatient to an outpatient setting, a part of the building may be closed and the staff size decreased. Still, for the purposes of most decisions in clinical practice, costs can be considered to be fixed or variable.

Productivity Costs

In contrast to direct costs, productivity costs do not stem from transactions for goods or services. Instead, they represent the cost of morbidity (e.g., time lost from work) or mortality (e.g., premature death leading to removal from the work force). They are costs because they represent the loss of opportunities to use a valuable resource, a life, in alternative ways. A variety of techniques are used to estimate productivity costs of illness or health care.^{23–27} Sometimes, as with patients infected with

human immunodeficiency virus,²⁸ the productivity costs of an illness are substantially greater than the direct costs of the illness.

Intangible Costs

Intangible costs are those of pain, suffering, and grief. These costs result from medical illness itself and from the services used to treat the illness. They are difficult to measure as part of a pharmacoeconomic study, though they are clearly considered by clinicians and patients in considering potential alternative treatments. Although investigators are developing ways to measure intangible costs—such as willingness-to-pay analysis whereby patients are asked to place monetary values on intangible costs³—at present these costs are often omitted from clinical economics research.

Perspective of Analysis

The third axis in Figure 35.1 is that of the perspective of an economic analysis of medical care. Costs and benefits can be calculated with respect to society's, the patient's, the payer's, and the provider's point of view. A study's perspective determines how costs and benefits are measured, and the economist's strict definition of costs (the consumption of a resource that could otherwise be used for another purpose) no longer may be appropriate when perspectives different from that of society as a whole are used. The economic impact of an intervention will be reported differently depending upon the perspective taken.

For example, a hospital's cost of providing a service may be less than its charge. From the hospital's perspective, then, the charge could be an overstatement of the resources consumed for some services. However, if the patient has to pay the full charge, it is an accurate reflection of the cost of the service to the patient. Alternatively, if the hospital decreases its costs by discharging patients early, the hospital's costs may decrease, but patients' costs may increase because of the need for increased outpatient expenses that are not covered by their health insurance plan.

Similarly, the cost to society is the opportunity cost, the value of the opportunities foregone

because of the resource having been consumed. Society's perspective usually is taken by measuring the consumption of real resources, including the loss of potentially productive human lives. As already noted, this cost does not count transfer payments, such as social security benefits. (From the point of view of the Social Security Administration, however, these payments would be a cost, because the perspective of the Social Security Administration is not the perspective of society.)

Because the costs of medical care may not be borne solely by the same parties who stand to benefit from it, economic analysis of medical care often raises vexing ethical problems related to equity, distribution of resources, and responsibility for the health of society's members.^{29, 30}

In summary, economic analysis of medical technology or medical care evaluates a medical service by comparing its dollar cost with its dollar benefit (cost–benefit), by measuring its dollar cost in relation to its outcomes (cost–effectiveness), or simply by tabulating the costs involved (cost identification). Direct costs are generated as services are provided. In addition, productivity costs should be considered, especially in determining the benefit of a service that decreases morbidity or mortality. Finally, the perspective of the study determines the costs and benefits that will be quantified in the analysis, and sensitivity analyses test the effects of changes in variable specifications for estimated measures on the results of the study.

METHODOLOGIC ISSUES IN THE PHARMACOECONOMIC ASSESSMENT OF THERAPIES

The basic approach for performing economic assessments of pharmaceutical products, as discussed above, has been adapted from the general methodology for cost–effectiveness and cost–benefit analysis. These methods have been well developed in medical technology assessment as well as in other fields of economic research. However, there remain a number of methodological issues that confront investigators in economic evaluations of pharmaceutical therapies. This section reviews some of these issues as they arise

in the design, analysis, and interpretation of pharmacoeconomic evaluations.

Clinical Trials Versus Common Practice

One of the most vexing of these issues is how to assess the cost implications of products during clinical trials. Ascertaining whether or not a product's costs are adequately offset by its effects or benefits presents a number of issues for consideration.^{31,32} We shall discuss some of these issues related to the case of the pharmacoeconomic assessment of a new prophylactic therapy for thromboembolic disease.

The Problem

As has been pointed out in other chapters of this volume, 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 poses 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. For example, in a study of prophylaxis against thromboembolic events, the protocol may specify testing of all patients for deep vein thromboses (DVTs) (e.g., fibrinogen scanning, venograms, or Doppler testing), whether or not the patients show clinical signs of these events. While this diagnostic strategy may be appropriate, it is not necessarily common practice. Yet, it can have wide ranging effects on the calculated costs and outcomes of care.

First, the protocol may induce the detection of extra cases—cases that would have gone undetected if no protocol were used in the usual care of patients. These cases may be detected earlier than they would have been in usual care. In the prophylaxis example above, repeated testing of all patients is likely to increase the number of DVTs that are detected, especially if, in usual care, patients are only tested when they develop clinical symptoms or signs of DVTs. This extra or early

detection may also reduce the average costs for each case detected, because subclinical cases or those detected early may be less costly to treat than clinically detected cases. However, because these two potential biases—more cases, each of which may cost less—work in opposite directions, the total costs of care for the patients in the trial may or may not exceed those that would occur in usual care.

Second, protocol induced testing may lead to the detection of adverse drug effects that would otherwise have gone undetected. As above, the average costs of each may be less because the adverse effects would be milder. However, their frequency would obviously be higher, and they could result in additional testing and treatment.

Third, protocol induced testing also may lead to the occurrence of fewer adverse events from the pharmaceutical product than would occur in usual care. The extra tests done in compliance with the protocol may provide information that otherwise would not have been available to clinicians, allowing them to take steps to prevent adverse events and their resulting costs. For example, an antibiotic protocol may call for more frequent testing of creatinine levels than would be conducted in usual care. These tests may warn physicians of impending renal problems, allowing them to change the drug dosage or the antibiotic. Thus, cases of nephrotoxicity that would have occurred in usual care may be avoided. This potential bias of reducing the costs of side-effects and adverse events would tend to lower the overall costs of care observed in the trial compared to usual care.

Fourth, due to ethical obligations that arise when patients are enrolled in trials, outcomes detected in trials may be treated more aggressively than they would be in usual care. In trials, it is likely that physicians will treat all detected treatable clinical outcomes. In usual care, physicians may treat only those outcomes that in their judgment are clinically relevant. This potential bias would tend to increase the costs of care observed in the trial compared to usual care.

Fifth, protocol induced testing to determine the efficacy of a product or to monitor the occurrence of all side-effects, whether clinically detectable or

not, generally will increase the costs of diagnostic testing in the trial, because many of these tests likely would be omitted in usual care. Alternatively, the protocol may reduce these costs in environments where there is overuse of testing. In teaching settings, for example, some residents may normally order more tests than are needed, and this excess testing may be limited by the protocol's testing prescriptions.

Sixth, clinical protocols may offer patients additional resources that are not routinely available in clinical practice. These additional resources may provide health benefits to patients. For example, protocols offering extensive home care services may affect the observed benefits of a therapy if the nursing intervention improves the management of the patient's illness. This could result in a bias in the study design if there are differences in the amount of home care services provided to patients in the treatment and control arms of a trial, or may result in additional health benefits to all study patients.

Seventh, patients in trials often are carefully selected. If a study sample has a mean patient age of 45, the result of the trial may not be readily generalizable to substantially older or younger populations. Similarly, exclusion criteria in clinical protocols may rule out patients with specific clinical syndromes (e.g., diabetes mellitus), women of childbearing potential, or patients of advanced age. These patients may require additional resources or may receive less benefit from therapy because their lifespan is shorter. These exclusions further limit the generalizability of the findings of efficacy studies.

A related issue in pharmacoeconomics trials is the generalizability of the health care delivery system of the patients in the study. A pharmacoeconomic study conducted through an HMO using its members as subjects may observe fewer referrals to specialist physicians than would the same clinical study in a different practice setting. This effect may be even more pronounced in multinational clinical trials, where health care systems, physician education, and patients' expectations for treatment differ by country.

Other difficulties in projecting the results of clinical trials to usual care arise because the

patients in clinical trials generally comply more completely with their treatment than do patients in usual care; they receive prescribed patterns of care; and because trials often have a placebo arm. If there is an actual placebo effect, this last factor may tend to understate the effectiveness the agent will have when it is utilized in usual care.

Routinely appending economic evaluations to clinical trials will likely yield "cost–efficacy" analyses, the results of which may be substantially different from the result of cost–effectiveness analyses conducted in the usual care setting. The problem of generalizability is similar to that found in clinical epidemiology research. However, clinical economics explicitly recognizes the added complexity of having different resource induced costs and benefits derived from clinical protocols and from observing patients in different health care systems in multicenter clinical trials.

Possible Solutions

One possible solution to this problem will be illustrated by examining the impact of a "usual care" arm appended as a third arm of a clinical trial. In such a three-arm study, patients randomized to the usual care arm of the study would be treated as they would be outside of the trial, rather than as mandated by the study protocol, and economic and outcomes data from usual care could thus be collected. These data would make it possible to quantify the number of outcomes that likely would be detected in usual care and the costs of these outcomes.

One drawback to this method is that physicians in the trial may treat all patients similarly, whether they are in the protocol driven arm or the usual care arm of the study. This contamination can be partially overcome by randomizing physicians to the protocol or usual care arms, and can be overcome more completely by randomizing the sites of care (e.g., different hospitals for different arms of the study). However, these options require large numbers of physicians and/or sites of care and, thus, are very costly to implement.

A second method that has been used to overcome these problems⁷ is to collect data retrospectively

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, to insure comparability between the data for usual care and trial patients, the data for trial patients may have to be collected retrospectively as well.

Two problems arise when using a concurrent comparison group to project the results of a trial to usual care. First, as with the randomization scheme above, the use of a protocol in the trial may affect the care delivered to patients who are not in the trial. If so, usual care patients may not receive the same care they would have received if the trial had not been performed. Thus, the results of the trial may lose generalizability to other settings. Second, the trial may enroll a particular type of patient (e.g., investigators may “cream-skim” by enrolling the healthiest patients with the fewest complications), possibly leaving a biased sample (e.g., of sicker and more complicated patients) for inclusion in the concurrent comparison group. This potential bias would tend to affect the estimate of the treatment costs that would be experienced in usual care.

Adoption of a historical comparison group would offset the issue of contamination. Because the trial was not ongoing when these patients received their care, it could not affect how they were treated. A historical comparison group would also tend to offset the selection bias: the subset of patients who would have been included in the trial if it had been carried out in the historic period will be candidates for the comparison group. However, use of a historic comparison group is unlikely to offset this bias entirely. Because this group is identified retrospectively, its attributes likely will reflect those of the average patients eligible for the trial, rather than those of the subset of patients that would have been enrolled in the trial (e.g., if cream-skimming had occurred).

However, differences between the care provided to patients in the trial and that provided to

patients in this group may be due as much to secular trends in the provision of medical care as they are to the adoption of a study protocol. For example, length of stay in the United States has decreased since the early 1980s, due in part to the implementation of the Medicare Prospective Payment System. Thus, historical cohorts from earlier periods may have had longer lengths of stay as inpatients than is currently seen in clinical practice. These data may suggest a protocol induced decrease in length of stay when one really does not exist.

To avoid these difficulties, the usual care comparison group may include both historic and concurrent comparison groups. In this case, multivariable methods such as multiple regression analysis or other analytic techniques must be used to control for differences among the historic and concurrent comparison groups as well as between the comparison groups and the patients in the trial. For example, in a regression analysis of length of stay in the trial and in usual care, variables representing each of the groups will indicate the magnitude of the secular trends, the selection bias, and the protocol effects of the trial.

A number of methods currently are being investigated to help overcome the potential biases of resource induced costs and benefits in clinical trials. These approaches include the development of “large and simple clinical trials,” increased attention to the generalizability of patient selection criteria in study design, and conducting the trial in different health systems simultaneously to assess the impact of the therapy in different delivery settings (e.g., using a large health maintenance organization as a clinical testing site).

Issues in the Design of Prospective Pharmacoeconomic Studies

We have already addressed some of the general issues in the design and interpretation of pharmacoeconomic studies. Yet, prospective pharmacoeconomic studies, especially within phase III clinical trials, are often our only opportunity to collect and analyze information on new therapeutic products

before decisions are made concerning insurance reimbursement and formulary inclusion for these agents. We now address issues that arise in the design of these studies.

Sample Size

The size required of the sample to identify a meaningful economic difference is frequently problematic. 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 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. Furthermore, in many cases the variances for the pharmacoeconomic data are unknown. 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 (see Table 35.1). More recently, several investigators have proposed methods for calculating the sample size for economic evaluations.³³

Participation of Patients

Those planning phase III clinical trials usually are more focused on the clinical results of the trial than they are on the economic results; they would usually like to keep the number of centers needed to complete the trial to a minimum; and they would rather finish the trial sooner than later. Thus, they have a concern that patients might agree to participate in the clinical trial, but not be willing to participate in the economic portion of the trial. In such a case, the investigators often argue that patients should be allowed to

Table 35.1. Study differences detectable given a fixed sample size

Standard deviation (LOS/\$)	Detectable difference R^2 for covariables			
	0.0	0.1	0.2	0.3
<i>n</i> = 150/group				
5	2	2	1	1
10	3	3	3	3
20	6	6	6	6
30	10	9	9	8
40	13	12	12	11
50	16	15	14	14
100	32	31	29	27
500	162	153	145	135
1000	324	307	289	271
2500	809	767	723	677
5000	1618	1535	1447	1354
<i>n</i> = 300/group				
5	1	1	1	1
10	2	2	2	2
20	4	4	4	4
30	7	7	6	6
40	9	9	8	8
50	11	11	10	10
100	23	22	20	19
500	114	109	102	96
1000	229	217	205	191
2500	572	543	512	479
5000	1144	1085	1024	957
<i>n</i> = 450/group				
5	1	1	1	1
10	2	2	2	2
20	4	4	3	3
30	6	5	5	5
40	7	7	7	6
50	9	9	8	8
100	19	18	17	16
500	93	89	84	78
1000	187	177	167	156
2500	467	443	418	391
5000	934	886	836	782

Note: Values represent minimum detectable differences between trial arms given the standard deviation reported for the row in the table, and a fixed sample size for each arm of the trial.

participate in the clinical portion of the trial but be excluded from the economic portion of the trial. While self-selection always poses difficulties for trials, it should be clear that this suggestion is particularly worrisome. The economic assessment would end up comparing an estimate of effects from the entire sample with an estimate of costs

from a nonrandom subset of the entire sample, thus allowing substantial bias to enter the analysis. Protocols should allow prospective collection of resource consumption and patient preference data, while sometimes incorporating a second consent to allow access to patients' financial information. This allows patients to consent to the clinical protocol and decline participation in the financial data collection, preventing the patient selection bias. However, given the low rates of refusal to the release of financial information, a single consent form should be considered for all trial data.

Data Collection

In many cases, by the time clinical investigators think to include economic assessments in their trials, they generally have asked for the collection of so much clinical data that it is nearly impossible to ask the data collectors to collect any economic data. Collection of resource consumption data from primary or secondary sources is essential for a prospective economic evaluation of a pharmaceutical therapy. Some data elements, such as patient preference assessments, can only be collected on a prospective basis. Other data elements, such as outpatient physician treatment records for a linked inpatient and outpatient economic evaluation of a therapy, or patient resource consumption information for many European hospitals without centralized billing systems, must be collected prospectively to simplify the data collection process for the study.

While some prospective data collection is required for almost all pharmacoeconomic studies, the amount of data to be collected for the pharmacoeconomic evaluation is still the subject of much debate. 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.

Resources are divided into specific categories for assessment for prospective data collection: inpatient resource use, outpatient resource use, and non-acute-care resource use. Within each of these categories, data can be subdivided into several categories: professional services (physicians, nurses, allied health professionals), hospital setting (intensive care unit, step-down unit, general medical floor), major diagnostic tests, (radiologic tests, laboratory tests, nuclear medicine studies), major surgical procedures (operations and non-operating-room procedures), and medications. Sample data collection forms for inpatient and outpatient resource consumption are presented as Figures 35.2 and 35.3. Issues related to data collection for economic studies have been reviewed recently.³³

Appropriate Comparators

Selection of appropriate treatment alternatives in a clinical study is essential for a useful economic evaluation of a pharmaceutical therapy. This issue is both a clinical and an economic one. Comparators can be the most common alternative therapies for a condition, or the lowest possible cost alternatives, even when not frequently used. However, in pharmacoeconomic studies, treatment comparators may be inappropriately selected as much for their relatively high price as they for their likely effectiveness. Phase III studies have special limitations in this regard, because agents will be compared against the placebo to assess efficacy rather than against alternative treatments to assess the relative effectiveness of the agent.

Multicenter Evaluations

Study results report an average cost-effectiveness ratio of a therapy given that the costs of study patients generally are not representative of any one of the centers in the trial. The argument here is that clinical results are about biology, which is

HOSPITAL DISCHARGE FORM		Patient No. _____
Principal Investigator _____	Study Hospital _____	Date of Admission ___/___/___
		Date of Discharge ___/___/___
Source of Admission <input type="checkbox"/> Emergency room <input type="checkbox"/> Transfer (from _____) <input type="checkbox"/> Elective	Discharge Diagnosis 1. _____ 2. _____	
Unit Type		Number of Days
Intensive Care Unit		
Intensive Care Unit with Mechanical Ventilator		
Bone Marrow Transplant Unit		
Step-Down/Intermediate Care Unit		
General Medical or Surgical Floor		
Pharmacologic Therapy		Total Dose
<input type="checkbox"/> Study Drug _____		
<input type="checkbox"/> Control _____		
Continuous IV Medication _____		
Types of Procedures		Date
Diagnostic Tests		Number of Tests
MRI		
CT Scan		
Bone Scan		

Figure 35.2. Inpatient resource assessment. This is a sample case report form for prospective assessment of inpatient resource consumption in a pharmaco-economic study.

OUTPATIENT VISIT RECORD		Patient No. _____				
Name of Physician and Location of Visit (eg., Emergency Room, Outpatient Clinic, Day Surgery, Home, Office)	Duration (in minutes)					
	date _/_	date _/_	date _/_	date _/_	date _/_	
	1.					
	2.					
Name of Nurse Clinician and Location of Visit (eg., Emergency Room, Outpatient Clinic, Day Surgery, Home, Office)	date _/_	date _/_	date _/_	date _/_	date _/_	
	1.					
	2.					
Type of Procedure		Date				
1.						
2.						
3.						
Diagnostic Tests		Number of Tests				
MRI						
CT Scan						
Bone Scan						
Other						
Other Therapy (medications, etc.)		Date				
1.						
2.						
3.						

Figure 35.3. Outpatient resource assessment. This is a sample case report form for prospective assessment of outpatient resource consumption in a pharmacoeconomic study.

relatively unaffected by state or, increasingly, national boundaries. Economic results, on the other hand, are highly dependent upon the setting in which the care is delivered. This may be especially true for trials in which different centers are in different countries with different health care systems and different patterns of treatment for patients with defined clinical conditions.

One might conclude that economic results should not be combined across centers, given the potential differences in resource consumption and resource pricing. However, this argument rests on the assumption that the resources used in the care of a patient do not affect the outcome of therapy. We are unaware of any evidence that supports such an assumption. In fact, resources may contribute directly to the treatment outcome (e.g., additional monitoring at one center may discover adverse side effects of a new treatment in time to prevent a resource intensive complication from developing; home nursing visits may prevent additional hospitalizations for trial patients). To the extent that people are willing to conduct subanalyses of efficacy results, economists should be willing to conduct subanalyses of economic results. However, if the disaggregation of clinical data occurs only at very gross levels, such as by continent, then only continent-specific differences in cost should be reported.

Factors Affecting Resource Consumption

Pharmacoeconomic research holds as a basic assumption the proposition that clinical severity of disease is the sole determinant of resource use by patients. Studies of regional variation, such as those by Wennberg *et al.*,³⁴ highlight the shortcomings of this assumption. This creates a significant challenge for health services research, and for pharmacoeconomics in particular. For example, when a new therapy is introduced to reduce severity of disease as a substitute for physician services that similarly reduce the severity of disease, if physicians either continue to provide the service to maintain their clinical practice or change the characteristics of the patients to whom they provide services (i.e.,

operate on less severely ill patients), we will not achieve the potential economic advantage afforded by the new therapy.

Economic Data

Analysts generally have access to resource utilization data such as length of stay, monitoring tests performed, and pharmaceutical agents received. When evaluating a therapy from a perspective that requires cost data rather than charge data, however, it may be difficult to translate these resources into costs. For example, does a technology that frees up nursing time reduce costs, or are nursing costs fixed in the sense that the technology is likely to have little or no effect on the hospital payroll? Economists taking the social perspective would argue that real resource consumption has decreased and thus nursing is a variable cost. Accountants or others taking the hospital perspective might argue that, unless the change affects overall staffing or the need for overtime, it is not a saving. This issue depends in part on the temporal perspective taken by the analyst. In the short term, it is unlikely that nursing savings are recouped; in the long term, however, there probably will be a redirection of services. This analysis may also be confounded by the potential increase in the quality of care that nurses with more time may be able to provide to their patients. In countries that have a shortage of hospital beds, hospital administrators often do not recognize staffing savings from early discharge programs, because the bed will be occupied by a new patient as soon as the old patient is discharged.

Perspective

When perspectives other than the societal perspective are adopted, it is unclear what benefits or outcomes should be counted in the analysis. For example, if a governmental agency's perspective is adopted, in which transfer payments such as pensions are counted as costs, quick deaths at age 65 may be valued more than long, costly deaths at age 75. Independent of whether we should condone this perspective, we must determine whether health

status is an independent goal to be included in the analysis.

In summary, due to their focus on efficacy and their use of clinical protocols, economic assessments of pharmaceutical products based upon phase III clinical trials are not without their problems. However, these issues can be developed in pharmacoeconomic analysis plans and addressed prospectively or through supplemental data collection activities conducted concurrently with the clinical trial.

Measurement and Modeling in Clinical Trials

Previously, we have discussed the development of pharmacoeconomic data throughout the drug development process. However, the types of datum 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, that is common to several economic trials, which allows for comparisons of economic data across clinical studies and is of relevance to policymakers.

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 following:

- the use of changes in blood cholesterol levels to demonstrate the efficacy of new lipid lowering agents (intermediate endpoint, changes in low density and high density lipoprotein levels; final endpoint, changes in myocardial infarction rate and survival; demonstration of the relationship

between intermediate and final endpoints, Framingham Heart Study);³⁵

- the use of change in blood pressure to demonstrate the efficacy of new antihypertensive agents (intermediate endpoint, changes in systolic and diastolic blood pressure; final endpoint, changes in stroke rates and survival; demonstration of the relationship between intermediate and final endpoints, Framingham Heart Study);³⁶ and
- the use of changes in CD4+ cells to demonstrate the efficacy of new therapies for patients infected with human immunodeficiency virus (intermediate endpoint, changes in CD4+ cells; final endpoint, changes in progression of disease and survival; demonstration of the relationship between intermediate and final endpoints, Walter Reed Army Medical Center HIV Staging System).³⁷

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.

The importance of this effort is illustrated in the following hypothetical example. A new therapy is under development that reduces the absolute risk of dying from a chronic disease by 50% as measured in a one-year trial. However, this therapy is not curative. A four-year trial was initiated at the same time as the one-year trial. The first-year results were the same in both the four-year trial and the one-year trial. However, there was an increased risk of death for treatment patients in the second and third years of the four-year trial, and by the end of the third year of the trial the survival rate was identical in the treatment and control arms of the four-year trial.

While there was a clear benefit to the new therapy in terms of postponing events from the first year of treatment to later years, the economic assessment of the therapy would suggest a greatly reduced treatment benefit from the four-year trial as compared with the one-year trial.

In projecting results of short term trials over patients' lifetimes, we usually present at least two of the many potential projections of lifetime treatment benefit.³⁸ A one-time effect model assumes that the clinical benefit observed in the trial is the only clinical benefit received by patients. Under this model, after the trial has ended, the conditional probability of disease progression for patients is the same in both arms of the trial. Given that it is unlikely that a therapy will lose all benefits as soon as one stops measuring them, this projection method generally is pessimistic compared to the actual outcome. A continuous-benefit effect model assumes that the clinical benefit observed in the trial is continued throughout the patients' lifetimes. Under this model, the conditional probability of disease progression for treatment and control patients continues at the same rate as that measured in the clinical trial. In contrast to the one-time model, this projection of treatment benefit most likely is optimistic compared to the treatment outcome.

While we and others have developed models as secondary analyses of new therapies,^{17, 18, 38-41} a number of clinical trials are now under way in which primary economic data are being collected. This change has resulted from an increasing awareness of the need for reliable economic data about new therapies at the time when the therapies are being introduced to the market. This impetus has also resulted from issues related to the complexity and cost of developing appropriate economic data for a secondary analysis of a new therapy, and issues related to the potential for bias in the design of economic studies conducted from analysis of secondary data sources.^{15, 16, 42-44} However, as illustrated above, even primary data collection in clinical trials does not eliminate the need for treatment models in the economic analysis of new therapies.

Analysis Plan For Cost Data

Analysis of cost data shares many features with analysis of clinical data. One of the most important is the need to develop an analysis plan before performing the analysis. Table 35.2 identifies a set of tasks that should be addressed in such a plan. The analysis plan should describe the study design (e.g., report on whether the trial is randomized and double blind; identify the randomization groups; outline the recruitment strategy; describe the criteria for patient evaluation) and any implications the design has for the analysis of costs (e.g., how one will account for recruiting strategies such as rolling admission and a fixed stopping date?).

The analysis plan should also specify the hypothesis and objectives of the study, define the primary and secondary 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 (e.g., costs and clinical outcomes at six months might be the primary outcome, while costs and clinical outcomes at 12 months might be a secondary outcome).

Also, the analysis plan should identify the statistical methods that will be used and how hypotheses will be tested (e.g., a *p*-value cutoff or a confidence interval for the difference that excludes zero). Further, the plan should prespecify whether

Table 35.2. 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 analysis
 8. Hypothesis tests
 9. Interim analyses
 10. Multiplicity issues
 11. Subgroup analysis
 12. Power/sample size calculations
-

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 (i.e., fully blinded rather than being simply blinded to treatment A versus treatment B). 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). Recently, a growing number of other methods have been used for making such determinations.

Univariate Analysis

A basic assumption underlying *t* tests and ANOVA (which are parametric tests) is that cost data are normally distributed. Given that the distribution of these data often violates this assumption, a number of analysts have begun using nonparametric tests, such as the Wilcoxon rank-sum test (a test of median costs) and the Kolmogorov–Smirnov test (a test for differences in cost distributions), which make no assumptions about the underlying distribution of costs. It has been reported that the Wilcoxon rank-sum test has 96% efficiency compared with the *t* test when the distribution is normally distributed, and that it is more powerful than the *t* test when data are not normally distributed.⁷

Table 35.3 shows the results of the univariate analysis of hospital costs measured among men receiving vehicle and an investigational medication for the treatment of aneurysmal subarachnoid hemorrhage.⁴⁵ The mean cost for patients receiving vehicle was \$20 287 (SD, \$22 542); the mean cost for patients receiving the investigational medication was \$25 185 (SD, \$22 619). The distribution (as seen from the quantiles reported in Table 35.3, which shows the distribution of costs for the two groups) is skewed. For example, the difference between the 25th and 50th percentiles is

Table 35.3. Hospital costs of tirilazad mesylate for subarachnoid hemorrhage in men

Variable	Vehicle	Tirilazad, 6 mg kg ⁻¹ per day
Cost (\$)	20 287	25 185
Standard deviation	(22 542)	(22 619)
Distribution		
5%	4 506	10 490
25%	9 691	13 765
50%	13 773	18 834
75%	23 044	31 069
95%	53 728	51 771
Comparison of differences		
<i>t</i> test	0.15	
<i>t</i> test (log of costs)	0.02	
Wilcoxon rank sum	0.001	
Kolmogorov–Smirnov	0.001	

approximately \$4500 for the two treatment groups, but it is approximately \$10 000 between the 50th and 75th percentiles. Of note, from the 5th to the 75th percentile, there was approximately a \$5000 difference between the two treatment groups. By the 95th percentile, the costs in the two groups were similar. These distributions provide evidence that the costs differ between the two treatment groups.

The parametric and nonparametric statistical tests, however, yielded conflicting conclusions about whether or not the cost differences were statistically different from one another. The *t* test comparing mean costs between the groups indicated a nonsignificant difference ($p = 0.15$), whereas the *t* test comparing the mean log of costs and both of the nonparametric statistical tests indicated they differed ($p < 0.02$). In this case, given the nonnormality of the data, it is likely that the 0.15 finding conceals the fact that there are differences in the costs of the two groups, and that the test of the log of costs and the nonparametric tests correctly indicate that costs differed significantly between the groups. Similarly conflicting conclusions about the statistical significance of observed differences in costs have been reported in other studies.⁴⁶

Use of statistical tests of the log of costs or of the median of costs to determine whether costs differ between the treatment groups may lead to a potential confusion about the outcome that is important for the analysis of the value for the cost of the new therapy (e.g., the cost-effectiveness ratio). Regardless of the statistical test used in this determination, the outcome used in the numerator of the cost-effectiveness ratio should always be the difference in mean costs. In other words, if for technical reasons one statistically assesses differences in the log of costs or differences in medians to determine whether costs differ between two treatment groups, the outcome of interest should still be the difference in mean costs and should not be the difference in the log or median of costs.

Multivariable Analysis

Regression analysis often is used to assess differences in costs, in part because the sample size

needed to detect economic differences may be larger than the sample needed to detect clinical differences (i.e., to overcome power problems). Traditionally, ordinary least squares regression has been used to predict costs (or their log) as a function of the treatment group while controlling for covariables such as disease severity, costs prior to randomization, etc. However, use of the log of costs as the outcome variable simply to avoid statistical problems posed by untransformed costs leaves one with the problem that we are not interested in this outcome itself; rather we are interested in the difference in untransformed costs. In addition, the retransformation of the predicted difference in the log of costs into an estimate of the predicted difference in costs is not trivial.^{47,48}

While univariate *t* tests and ANOVAs assume the normal distribution of cost data, ordinary least squares regression assumes that the error terms from the prediction of costs are normally distributed. Because of the potential violation of this assumption, however, a number of alternative multivariable methods have recently been proposed for analyzing costs. These include nonparametric hazards models,⁴⁹⁻⁵³ parametric failure-time models,⁴⁹ Cox semiparametric regression,⁵⁴ and joint distributions of survival and cost.⁵⁵ The relative merits of several of these methods have been compared by Lipscomb and colleagues;⁵⁶ however, there is little conclusive evidence regarding which model is best in a given analytic circumstance.

Table 35.4 shows selected results of an ordinary least squares regression predicting hospital costs measured among men receiving vehicle and the investigational medication for the treatment of aneurysmal subarachnoid hemorrhage. On average, costs among those receiving the investigational medication were \$6058 higher than costs among patients receiving vehicle ($p = 0.03$). Increasing levels in the neurograde of subarachnoid hemorrhage upon entry to the study (grades of subarachnoid hemorrhage range from I to V, with V being the most severe) were generally associated with increasing costs; the reduction in costs among those in grade V was due principally to the large number of patients in this category who died in the hospital. Other predictors of hospital costs

Table 35.4. Selected coefficients and p values for the hospital cost regressions for men receiving tirilzad for subarachnoid hemorrhage

	Coefficient	p
Intercept	1747	0.90
Randomization group ^a		0.05
6 mg kg ⁻¹ per day	6058	
2 mg kg ⁻¹ per day	-100	
0.6 mg kg ⁻¹ per day	-247	
Neurograde of subarachnoid hemorrhage		0.0001
Grade 2	3950	
Grade 3	3904	
Grade 4	9132	
Grade 5	5406	

^a 6 mg kg⁻¹/day versus vehicle, 2 mg kg⁻¹/day, and 0.6 mg kg⁻¹/day, $p = 0.03, 0.03,$ and $0.02,$ respectively; no other comparisons statistically significant.

included the additional days between onset of subarachnoid hemorrhage and randomization into the trial (+), age (+), and country (+/-) (data not shown).⁴⁵

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 and for the resulting cost-effectiveness ratio. Recently a substantial literature has developed related to construction of confidence intervals for cost-effectiveness ratios.⁵⁷⁻⁶¹

One of the most dependably accurate methods for deriving 95% confidence intervals for cost-effectiveness ratios is the nonparametric bootstrap method.⁶¹ In this method, one resamples 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.⁶²

In the multivariable regression analysis above, we estimated that therapy with the investigational medication added \$6058 to the cost of hospitalization (95% CI, \$693-11 423). The results of a logistic regression predicting death indicated that the investigational medication yielded a difference in the predicted probability of death of 0.225.⁴⁵ The cost per death averted was \$26 924 (\$6058/0.225). The results of the bootstrap analysis indicated that the 95% CI for the cost-effectiveness ratio ranged from \$4300 to \$54 600.⁴⁸ Interpreting the results of the bootstrap in a Bayesian sense, evaluating stochastic uncertainty alone, there is a 96% chance that the ratio is below \$50 000 per death averted.

In addition to addressing stochastic uncertainty, one may want to address uncertainty related to parameters measured without variation (e.g., unit cost estimates, discount rates, etc.), whether or not the results are generalizable to settings other than those studied in the trial, and, for chronic therapies, whether the cost-effectiveness ratio observed within the trial is likely to be representative of the ratio that would have been observed if the trial had been conducted for a longer period. These sources of uncertainty are often addressed using sensitivity analysis.

CURRENTLY AVAILABLE SOLUTIONS

The previous sections of this chapter dealt with the principles of clinical economics and methodological issues surrounding the economic analysis of pharmaceutical products. This section presents a

set of case studies that illustrate the practical application of these methods to the evaluation of pharmaceuticals. The following cases illustrate cost-effectiveness analyses of granulocyte-macrophage colony stimulating factor (GM-CSF) and interleukin three (IL-3) for Hodgkin's and non-Hodgkin's lymphoma patients undergoing autologous bone marrow transplantation (ABMT), tirilazad mesylate for aneurysmal subarachnoid hemorrhage, and epoprostenol for patients with severe congestive heart failure.

COST-EFFECTIVENESS OF IL-3/GM-CSF FOR HODGKIN'S AND NON-HODGKIN'S LYMPHOMA PATIENTS UNDERGOING AUTOLOGOUS BONE MARROW TRANSPLANTATION

In this study, cost-effectiveness analysis was used to assess the economic effect of a new cytokine therapy, IL-3 with GM-CSF, compared to the standard care of GM-CSF alone in patients undergoing autologous bone marrow transplantation (ABMT) for treatment of Hodgkin's and non-Hodgkin's lymphoma.⁶³ Patients enrolled in the study were randomized to receive either 21 days of GM-CSF starting on the day of ABMT or ten days of IL-3 starting on the day of ABMT followed by 11 days of GM-CSF. Patients were followed up to 12 months after treatment to assess safety. A total of 206 patients were enrolled in the trial from June 1993 through March 1995. The preliminary results of the phase III clinical trial indicated that there was no difference in hematopoietic recovery between the administration of IL-3/GM-CSF after ABMT compared to the administration of GM-CSF alone. An economic protocol ran parallel to the clinical trial.

Patients were enrolled in the economic study on a rolling admissions basis. An admissions stopping date was fixed at six months after the last patient was enrolled in the clinical protocol. A total of 115 of the 206 patients from the clinical arm of the trial were enrolled in the economic study.

Data were collected through monthly telephone interviews with the patient. Data collection included assessment of the initial hospital length of

stay, post-discharge resource consumption, and assessment of patients' health-related quality of life at baseline, three months, six months, nine months, and one year. Patients' quality of life was measured using the EuroQol EQ-5D thermometer, a visual analog scale. This scale measures a patient's preferences for their health state on a scale of zero to 100. A score of zero represented the worst imaginable health state, and 100 represented the best imaginable health state.

Costs were estimated for each of the resource categories assessed (rehospitalization, chemotherapy, radiation therapy, transfusion, outpatient procedures, and professional services). All costs are reported in 1995 dollars.

Hospital bills from the ABMT hospitalization and rehospitalizations provided information on charges for resource use during the hospital stay. These charges were converted into costs using hospital-wide cost-to-charge ratios from the Medicare cost report data set. Cost information was used to allow for better assessment of the actual use of resources by patients. Ordinary least squares regression was used to predict costs if complete hospital bills from the ABMT hospitalization or rehospitalizations were unavailable. Current Procedural Terminology (CPT) codes were used to identify physician services and were assigned costs using the 1995 Medicare fee schedule. The Medicare fee schedule was also used to assign costs to physician time for administering chemotherapy, radiation therapy, physician followup visits, and outpatient surgery and procedures. The costs for transfusions were based on cost data from a university hospital.

The cost for IL-3/GM-CSF was not included because therapy was part of the clinical protocol. The cost of chemotherapy was calculated by using the standard doses and frequencies for each patient's chemotherapy regimen. The cost was based on average wholesale price. Physician time investment for administration of the chemotherapy was also included in the cost of chemotherapy.

Whenever outpatient data were missing, costs were based on patient means when more than one month of data was available for that particular patient. If one month of data was unavailable, the mean cost across all patients for the month was

assigned to the patient with missing outpatient data.

The results of the resource use analysis indicate that there was no statistically significant difference between treatment groups in length of ABMT hospitalization or survival (Table 35.5). While not statistically significant, patients in the standard care group had more rehospitalizations than the IL-3/GM-CSF group. However, the average length of stay of rehospitalizations was shorter for the IL-3/GM-CSF group than the standard care group. Patients in the GM-CSF alone group had fewer outpatient visits per month, required less radiation therapy, less chemotherapy, and fewer transfusions after discharge from ABMT hospitalization than the IL-3/GM-CSF group. However, these differences were not statistically significant. The results of the cost analysis indicate that there were no statistically significant differences in the average ABMT hospitalization costs, survival weighted costs after discharge, total costs, or quality adjusted life months between the GM-CSF group and the IL-3/GM-CSF group. While not statistically significant, patients in the IL-3/GM-CSF group had higher outpatient costs per month than the GM-CSF group. However, patients in the IL-3/GM-CSF group had slightly lower rehospitalization costs than the GM-CSF group, but this was not statistically significant.

The authors conclude that there was no significant effect of IL-3 on costs of care for patients undergoing ABMT for the period up to 13 months after the procedure. The findings of this study are consistent with the model that IL-3 does not add any clinical benefit to GM-CSF. This study demonstrated the feasibility of prospective economic evaluation in phase III trials of new cancer therapies.

ECONOMIC ANALYSIS OF TIRILAZAD MESYLATE FOR ANEURYSMAL SUBARACHNOID HEMORRHAGE

The investigators undertook a cost-effectiveness analysis on the use of tirilazad mesylate, a potential free radical scavenger and lipid peroxidation inhibitor, for the treatment of aneurysmal subarachnoid hemorrhage in a randomized, double-blind, placebo controlled phase III clinical trial.⁴⁵ A sample of 1023 patients from nine European countries, Australia, and New Zealand were randomized to receive one of four treatments: vehicle or tirilazad 0.6, 2.0, or 6 mg/kg of body weight per day for eight to ten days. All treatments were administered within 48 hours of the occurrence of a subarachnoid hemorrhage and ceased ten days after the initial hemorrhagic event. All patients received nimodipine treatment concomitantly.

Clinical and economic outcomes at three months and hospital costs were estimated using data from 1019 of the 1023 patients enrolled in the study. Death during the three months after randomization was the clinical outcome. The primary clinical outcome was occurrence of vasospasm, and the secondary outcome was Glasgow Outcome Scale (GOS) score and mortality. The authors evaluated the costs of hospitalizations during the trial as well as patients' residence and employment status three months after randomization. Costs estimates were based on resource use and unit costs for resources used.

Data on the length of stay, number of imaging studies, number and types of surgical procedures, and medication use were collected prospectively. Information on the site of care in the hospital was obtained retrospectively but was collected before

Table 35.5. Results table from IL-3 manuscript

	Length of stay for randomization hospitalization (SD) ^a	Length of stay after discharge for ABMT (SD) ^b	Number of rehospitalizations
GM-CSF	31.97 (10.42)	4.46 (8.72)	49
GM-CSF/IL-3	33.47 (8.58)	4.45 (8.49)	33

^ap value for difference in hospital stays = 0.397.

^bp value for difference in hospital stays = 0.913. Among patients with rehospitalizations, the average hospital stay was 5.18 days and 7.15 days ($p = 0.16$) for GM-CSF and GM-CSF/IL-3, respectively.

the study results were known and while the investigators were blinded to the treatment groups. Patient status at three months was evaluated prospectively by assessing daily residence costs at three months for patients living at home with supervision or dependent on others as well as for those in minimal care, skilled care, or long term rehabilitation institutions. The daily employment value at three months was also assessed for homemakers and for full- and part-time workers.

Local health economists from six countries collected unit costs of inpatient resource utilization. The averages of the unit costs from the six countries were used for the five other countries. The \$137.50 cost per 150 mg of tirilazad was based on a price set by the manufacturer. The authors determined values for employment using wage and salary data from the participating countries. The unit costs from other countries were converted into 1993 US dollars. The authors state that during the sensitivity analysis, deaths averted were translated into gains in life expectancy both with and without adjustments for quality of life.

The results of the study indicate that patients were similar in all groups except in the proportions having right-to-left and left-to-right shifts of the midline structures and those having generalized, as opposed to localized, brain swelling. Total length of hospital stay, number of days between the onset of subarachnoid hemorrhage and randomization, number of days the patient was intubated, characteristics of the hemorrhage, the country in which patients received care, and mortality of the patients were all predictors of stay by unit type.

Results of the economic analysis showed that the average hospital cost was \$20 341 (SD, \pm \$17 239) for the whole sample. The average hospital cost for women (\$19 569 \pm \$15 156) was less than the cost among men (\$21 835 \pm \$20 743). The results also indicated that the majority of the cost was attributable to length of stay and the most difference in cost was due to the costs of tirilazad. The cost analysis at three months showed that the largest difference in employment value was observed between men who received tirilazad 6 mg kg⁻¹ per day and those who received vehicle (\$9.20 additional earnings per day). In addition, the results showed that the largest difference in

residence cost was also between these two groups (\$15.80 additional residence cost per day). However, none of these differences was statistically significant. One significant finding of this study was that those who received tirilazad 6 mg kg⁻¹ per day had a significant reduction in the probability of death in the whole sample ($p = 0.002$) and in men ($p = 0.0001$). There was no significant difference in the probability of death among women between the group who received tirilazad 6 mg kg⁻¹ per day and those who received vehicle. When costs and outcomes were compared, the results showed that in both the entire sample and in men, tirilazad 6 mg kg⁻¹ per day was associated with improved survival compared to vehicle, but also to increased hospital costs. The cost per death averted was \$29 615 for the sample as a whole and \$26 924 for men. There were no significant differences in costs or probability of survival of women in either the tirilazad 6 mg kg⁻¹ per day or vehicle group.

The results were subjected to a sensitivity analysis, showing that the cost-effectiveness ratio (95% confidence interval) between those in the entire sample who received tirilazad 6 mg kg⁻¹ per day and vehicle was \$9189 per death averted due to tirilazad, adding hospital costs and mortality. The cost-effectiveness ratios among men (95% CI) ranged from \$4300 to \$54 600 per death averted. The sensitivity analysis also showed that in 68.8% of women, 6 mg kg⁻¹ per day of tirilazad resulted in an increase in hospital costs and survival. Five percent experienced decreased costs and survival, 11.6% had decreased costs and increased survival, and 14.3% had increased costs and decreased survival. Another finding was that in the entire sample, the ratios of cost per year of life saved and cost per quality adjusted year of life saved fell below \$50 000 if survivors live on average 0.6 and 0.8 years respectively. For men, these ratios fell below \$50 000 if survivors at the end of the trial lived an average of 1.1 and 2.4 years. Among men, the ratio of cost per year of life saved did not fall below \$27 500. Also, the ratio of the cost per quality adjusted year of life saved did not fall under \$36 400.

The economic analysis of this study showed that treatment with tirilazad mesylate is associated with

a significant increase in survival and increase in the cost of care. The results also showed that the ratios of cost per death averted, cost per year of life saved, and cost per quality adjusted year of life saved are favorable when compared to other interventions.

RESULTS OF THE ECONOMIC EVALUATION OF THE FIRST STUDY: A MULTINATIONAL PROSPECTIVE ECONOMIC EVALUATION

The Flolan International Randomized Survival Trial (FIRST) evaluated the use of epoprostenol (Flolan), a potent vasodilator, for the treatment of patients with severe congestive heart failure. The economic evaluation was conducted concurrently with the phase III clinical trial.⁵³ The primary objective of the economic evaluation was to assess whether additional resources required for the use of epoprostenol delivered as a continuous infusion were offset by reductions in other resources required in patient care, or were related to health benefits sufficient to justify the use of this therapy in heart failure patients. The secondary objective was to assess the effect of epoprostenol on quality of life. The clinical trial ended early due to increased mortality in patients receiving epoprostenol.

Two economic analyses were conducted: the first assessed the costs and outcomes of patients randomized to epoprostenol with best usual care or best usual care alone; the second analysis projected costs and outcomes that would have been observed had all patients been followed for 12 months.

Data on resource consumption, unit costs of resources, and patient quality of life were collected. Case report forms were used to collect data on inpatient hospital days, inpatient procedures, in- and outpatient physician and nursing visits, outpatient laboratory use, and nonacute care (i.e., nursing home) required for care of the patient. Total costs of care were calculated by multiplying the unit costs of resource categories by the counts of medical services. Unit costs for inpatient services were drawn from a cost accounting system in a single university hospital in the United States.

Unit costs for physician services were based on the 1994 Medicare Fee Schedule. Epoprostenol therapy was not included in the cost analysis because there was no market price available for the therapy. Both the EuroQol instrument (described above) and the Nottingham Health Profile (NHP), a generic measure of patients' subjective health status, were used to measure patient quality of life. The evaluation of quality adjusted life months (QALM) was based on the EuroQol instrument. QALMs were calculated by multiplying patients' survival time by their EuroQol scores. QALMs are expressed along a 0–1 scale.

A total of 471 patients were enrolled in the study, of which 232 were randomized to receive best usual care alone. At the time of randomization, both groups were similar except for the proportion of patients receiving assistance in daily living from someone other than a friend, relative, spouse, or home health aide ($p < 0.05$).

The results of the economic analysis showed that the cost of care for those patients receiving usual best care was \$11 797 (95% CI, \$9184 to \$14 412). The majority of this cost was accrued in the inpatient setting. The patients in this group had an average survival of 4.7 months (95% CI, 4.2 to 5.2) and 6.3 months of potential followup time (see Table 35.6). These patients also experienced 2.3 quality adjusted months of survival (95% CI, 2.0 to 2.6). In contrast, patients receiving epoprostenol had costs that were \$5022 more than the cost of

Table 35.6. Mean total costs in US dollars, survival time, and potential followup time among patients receiving epoprostenol and those receiving usual care (SD)

	Usual care	Epoprostenol
Costs (\$)		
Outpatient visits	1 174 (1 374)	1 331 (1 826)
Outpatient procedures	151 (207)	177 (339)
Inpatient stay	10 314 (19 994)	14 894 (25 871)
Inpatient procedures	158 (297)	417 (408)
Survival time (months)	4.7 (3.6)	4.3 (3.4)
Potential followup time (months)	6.3 (4.0)	6.2 (4.0)

the usual care group (95% CI, \$13 453 to \$20 187), a statistically significant difference. Similar to the usual care group, most of the costs were incurred in the inpatient setting. The epoprostenol patients survived an average of 4.3 months (95% CI, 3.9 to 4.7). However, this was not statistically different from the usual care group. The epoprostenol group also experienced a decrease in the quality adjusted months of survival (2.2 months, 95% CI, 1.9 to 2.5) but this was not statistically significant. Ordinary least squares regression was used to predict the total costs during the trial. The results show that epoprostenol was associated with increased cost of care for patients ($p < 0.05$). Duration of potential followup time and requirement of assistance with activities of daily living at the time of randomization were associated with increased costs or a trend toward increased costs. Unemployment status and chronic hypertension at randomization were associated with decreased costs or a trend toward decreased costs. An increase in utility adjusted survival was associated with the EuroQol score at randomization, duration of potential followup time, disabled status at randomization, and greater than 60 ft min⁻¹ in the 6 min walk test. New York Heart Association (NYHA) class IV heart failure at randomization was associated with a decrease in quality adjusted survival. However, this only trended toward significance ($p < 0.10$). There was no significant association of epoprostenol with a decrease in survival time.

The results of the analysis on expected 12-month costs and survival time showed that the expected treatment costs during the year for the usual care group were \$22 476 (95% CI, \$15 268 to \$29 863). The costs declined during the first four months and then remained constant for the remainder of the study. The projected overall survival time and quality adjusted survival time for the usual care group were 7.1 (95% CI, 6.4 to 7.8) and 4.4 (95% CI, 3.8 to 4.9) months. The epoprostenol group had costs of \$27 747 (95% CI, \$21 806 to \$34 228). Their costs were higher than the usual care group for the seven months after randomization and then became lower than the cost for the usual care group. The projected overall survival

and quality adjusted survival for the epoprostenol group were 5.8 (95% CI, 5.07 to 6.52) and 3.7 (95% CI, 3.08 to 4.23) months. In the epoprostenol group, there was a difference in survival of -1.3 months (95% CI, -1.27 to -1.34) and in quality adjusted survival of -0.67 months (95% CI, -0.65 to -0.70). The quality adjusted survival was greater in some months than in preceding months, which indicates that the improvements in quality of life offset decreases in survival in those months.

The authors concluded that this study successfully quantifies resource utilization and quality of life in a population with severe heart failure and in explaining the variation between patients receiving epoprostenol with usual best care and usual best care alone. This study demonstrates that an economic analysis can be conducted during a phase III clinical trial.

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 decisionmakers. 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.

The naive decisionmaker might weigh drugs according to their purchase price alone. This paradigm ignores two essential elements in choosing pharmaceuticals. First, in identifying a drug's cost, its purchase price is only part of its real economic impact. The costs of preparation and delivery, as well as the cost of monitoring for and

treating adverse events and side-effects, are unavoidable elements of the cost of treating patients.

Second, a full analysis should go beyond the identification of cost. Only if the safety and effectiveness of two pharmaceutical agents are equivalent will cost alone determine the choice of therapy. Cost-effectiveness analysis requires that cost be weighed against effectiveness and that when two or more alternatives are being compared, the additional cost per additional unit of effectiveness be measured. Beyond these considerations of cost identification and cost-effectiveness, a full economic analysis will also assess the net value, or utility, of the drug's clinical contribution.

As physicians are asked simultaneously to represent their patients' interests while being asked to deliver clinical services with parsimony, they increasingly will need to turn for assistance to collaborative efforts of epidemiologists and economists in the assessment of new therapeutic agents. Through a merger of epidemiology and economics,⁶⁴ better information can be provided to clinical decision makers, and limited resources can be used most effectively for the health of the public.

ACKNOWLEDGEMENT

This is to give special acknowledgement to Harris Koffer, PharmD, Corporate Vice President and General Manager of Covance, Inc., for his contributions to this chapter.

REFERENCES

- Bombardier C, Eisenberg J. Looking into the crystal ball: can we estimate the lifetime cost of rheumatoid arthritis? *J Rheumatol* 1985; **12**: 201-4.
- Drummond MF. *Principles of Economic Appraisal in Health Care*. New York: Oxford University Press, 1980.
- Drummond MF, Stoddart GL, Torrance GW. *Methods for the Evaluation of Health Care Programs*. New York: Oxford University Press, 1987.
- Drummond MF, Davies L. Economic analysis alongside clinical trials: revisiting the methodological issues. *Int J Technol Assess Health Care* 1991; **7**: 561-73.
- Detsky AS, Naglie IG. A clinician's guide to cost effectiveness analysis. *Ann Intern Med* 1990; **113**: 147-54.
- Eddy DM. Cost-effectiveness analysis: is it up to the task? *J Am Med Assoc* 1992; **267**: 3342-8.
- Eisenberg JM, Koffer H, Finkler SA. Economic analysis of a new drug: potential savings in hospital operating costs from the use of a once-daily regimen of a parenteral cephalosporin. *Rev Infect Dis* 1984; **6**: S909.
- Eisenberg JM, Koffer H, Glick HA, et al. What is the cost of nephrotoxicity associated with aminoglycosides? *Ann Intern Med* 1987; **107**: 900-9.
- Eisenberg JM. Clinical economics: a guide to the economic analysis of clinical practices. *J Am Med Assoc* 1989; **262**: 2879-86.
- Glick HA, Connell ML, Koffer H, et al. The economic impact of prophylaxis against venous thromboembolism in patients undergoing total hip replacement. *Clin Res* 1988; **36**: 337A.
- Warner KE, Luce BR. *Cost Benefit and Cost Effectiveness Analysis in Health Care: Principles, Practice and Potential*. Ann Arbor, MI: Health Administration Press, 1982.
- Weinstein MC. Economic assessments of medical practices and technologies. *Med Decis Making* 1981; **1**: 309.
- Weinstein MC, Stason WB. Cost-effectiveness of coronary artery bypass surgery. *Circulation* 1982; **66**: 56-68.
- Weinstein MC. Principles of cost-effective resource allocation in health care organizations. *Int J Technol Assess Health Care* 1990; **6**: 93-103.
- Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Can Med Assoc J* 1992; **146**: 473-81.
- Evans D, Freund D, Dittus R, Robertson J, Henry D. *The Use of Economic Analysis as a Basis for Inclusion of Pharmaceutical Products on the Pharmaceutical Benefits Scheme*. Canberra: Department of Health, Housing and Community Services, Commonwealth of Australia, 1990.
- Schulman KA, Kinoshian BP, Jacobson TA, Glick HA, Willian MK, Koffer H, Eisenberg JM. Reducing high blood cholesterol level with drugs: cost effectiveness of pharmacologic management. *J Am Med Assoc* 1990; **264**: 3025-33.
- Schulman KA, Glick HA, Rubin H, Eisenberg JM. Cost effectiveness of HA-1A monoclonal antibody for gram-negative sepsis: prospective economic assessment of a new therapeutic agent. *J Am Med Assoc* 1991; **226**: 3466-71.
- Finkler SA. The distinction between cost and charges. *Ann Intern Med* 1982; **96**: 102-9.

20. Bridges JM, Jacobs P. Obtaining estimates of marginal cost by DRG. *Healthcare Financial Management* 1986; 40–6.
21. Hsiao WC, Braun P, Becker ER, *et al.* An overview of the development and refinement of the resource-based relative value scale: the foundation for reform of U.S. physician payment. *Med Care* 1992; **30** (suppl): NS1–12.
22. Granneman TW, Brown RS, Pauly MV. Estimating hospital costs. *J Health Econ* 1986; **5**: 107–27.
23. Landefeld JS, Seskin EP. The economic value of life: linking theory to practice. *Am J Public Health* 1982; **72**: 555–65.
24. Rice DP, Hodgson TA. The value of human life revisited. *Am J Public Health* 1982; **72**: 536–8.
25. Thompson MS, Read JL, Liang M. Feasibility of willingness-to-pay measurement in chronic arthritis. *Med Decis Making* 1984; **4**: 195–215.
26. Mishan EJ. *Cost-Benefit Analysis*, 3rd edn. London: Allen and Unwin, 1992.
27. Shelling TC, Chase SB, ed. The life you save may be your own. In: *Problems in Public Expenditure Analysis*. Washington, DC: Brookings Institution, 1968.
28. Lynn LA, Schulman KA, Eisenberg JM. The pharmacoeconomics of HIV disease. *Pharmacoeconomics* 1992; **1**: 161–74.
29. Strosberg MA, Wiener JM, Baker R, Fein IA, eds. *Rationing America's Medical Care: The Oregon Plan and Beyond*. Washington, DC: Brookings Institution, 1992.
30. Sulmasy DP. Physicians, cost-control, and ethics. *Ann Intern Med* 1992; **116**: 920–6.
31. Drummond MF, Stoddart GL. Economic analysis and clinical trials. *Control Clin Trials* 1984; **5**: 115–28.
32. Weinstein MC. Methodologic considerations in planning clinical trials of cost effectiveness of magnetic resonance imaging. *Int J Technol Assess Health Care* 1985; **1**: 567–81.
33. Brown M, Glick HA, Harrell F, *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.
34. Perrin JM, Homer CJ, Berwick DM, Woolf AD, Freeman JL, Wennberg JE. Variations in rates of hospitalizations of children in three urban communities. *New Engl J Med* 1989; **320**: 1183–7.
35. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991; **121**: 293–8.
36. Stokes J III, Kannel WB, Wolf PA, Dagostino RB, Lupples LA. Blood pressure as a risk factor for cardiovascular disease. The Framingham Study—30 years of follow-up. *Hypertension* 1989; **13** (suppl): I13–8.
37. Redfield RR, Wright DC, Tramont EC. The Walter Reed staging classification of HTLV-3/LAV infection. *New Engl J Med* 1986; **314**: 131–2.
38. 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.
39. Glick H, Heyse JF, Thompson D, Epstein RS, Smith ME, Oster G. A model for evaluating the cost-effectiveness of cholesterol-lowering treatment. *Int J Technol Assess Health Care* 1992; **8**: 719–34.
40. Goldman L, Weinstein MC, Goldman PA, Williams LW. Cost-effectiveness of HMG-CoA reductase inhibition for primary and secondary prevention of coronary heart disease. *J Am Med Assoc* 1991; **265**: 1145–51.
41. Grover SA, Abrahamowicz M, Joseph L, Brewer C, Coupal L, Suissa S. The benefits of treating hyperlipidemia to prevent coronary heart disease: expected changes in life expectancy and morbidity. *J Am Med Assoc* 1992; **267**: 816–22.
42. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Med Care* 1992; **30**: 2–243.
43. Hillman AL, Eisenberg JM, Pauly MV, *et al.* Avoiding bias in the conduct and reporting of cost-effectiveness research sponsored by pharmaceutical companies. *New Engl J Med* 1991; **324**: 1362–5.
44. Udvarhelyi IS, Colditz GA, Ri A, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly? *Ann Intern Med* 1992; **116**: 238–44.
45. Glick H, Willke R, Polsky D, Llana T, Alves WM, Kassel N, Schulman KA. Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage: economic evaluation of a phase III clinical trial in Europe and Australia. *Int J Technol Assess Health Care* 1998; **14**: 145–60.
46. Zhou X-H, Melfi CA, Hui SL. Methods for comparison of cost data. *Ann Intern Med* 1997; **127**: 752–6.
47. Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. *J Health Econ* 1998; **17**: 283–95.
48. Duan N. Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc* 1983; **78**: 605–10.
49. Dudley RA, Harrell FE, Smith LR, Mark DB, Califf RM, Pryor DB, Glower D, Lipscomb J, Hlatky M. Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. *J Clin Epidemiol* 1993; **46**: 261–71.
50. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997; **53**: 113–28.

51. Fenn P, McGuire A, Phillips V, Backhouse M, Jones D. The analysis of censored treatment cost data in economic evaluation. *Med Care* 1995; **33**: 851–61.
52. Fenn P, McGuire A, Backhouse M, Jones D. Modeling programme costs in economic evaluation. *J Health Econ* 1996; **15**: 115–25.
53. Schulman KA, Buxton M, Glick H, Sculpher M, Guzman G, Kong J, Backhouse M, Mauskopf J, Bell L, Eisenberg JM, and the FIRST Investigators. Results of the economic evaluation of the FIRST study. *Int J Technol Assess Health Care* 1996; **12**: 698–713.
54. Harrell F, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; **15**: 361–87.
55. Lancaster T, Intrator O. Panel data with survival: hospitalization of HIV patients (Brown University Department of Economics Working Paper Series 95-36). 1995.
56. Libscomb J, Ancukiewicz M, Parmigiani G, Hasselblad V, Samsa G, Matchar DBJ. Predicting the cost of illness: a comparison of alternative models applied to stroke. *Med Decis Making* 1998; **18**: S39–56.
57. Heyse J, Cook J. *A New Measure of Cost Effectiveness in Comparative Clinical Economic Trials*. Boston, MA: American Statistical Association, 1992.
58. O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost effectiveness studies in health care. *Med Care* 1994; **32**: 150–63.
59. Willan AR, O'Brien BJ. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Econ* 1996; **5**: 297–305.
60. Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Stat Med* 1996; **15**: 1447–58.
61. 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.
62. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.
63. Schulman KA, Dorsainvil D, Yabroff KR, DiCesare J, Felser J, Eisenberg JM, Glick HA, and the IL-3 BMT Study Team. Prospective economic evaluation accompanying a trial of GM-CSF/IL-3 in patients undergoing autologous bone marrow transplantation for Hodgkin's and non-Hodgkin's lymphoma. *Bone Marrow Transplant* 1998; **21**: 607–14.
64. Eisenberg JM. From clinical epidemiology to clinical economics. *J Gen Intern Med* 1988; **3**: 299–300.

36

Using Quality of Life Measurements in Pharmacoepidemiology Research

GORDON H. GUYATT

Department of Medicine and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

ROMAN JAESCHKE

Department of Medicine, St Joseph's Hospital, Hamilton, 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 (β -blockers after myocardial infarction), rate of hospitalization (neuroleptic agents for schizophrenia), rate of disease occurrence (antihypertensives for strokes), or rate of disease recurrence (some form of chemotherapy after surgical cancer treatment). Alternatively, clinicians frequently rely on direct physiological 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.

Clinical investigators have recognized that there are other important aspects of the usefulness of the interventions which these epidemiological, physiological, or biochemical outcomes do not address.

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 conclusion reached when evaluating different outcomes may differ: physiological measurements may change without people feeling better,^{1,2} a drug may ameliorate symptoms without a measurable change in physiological function, or life prolongation may be achieved at the expense of unacceptable pain and suffering.³ The recognition of these patient oriented (versus disease oriented) areas of well-being led to the introduction of a technical term: health related quality of life (HRQL).

Quality of life, as it is often used, lacks focus and precision and, because it is an abstract concept, its definition has led to much debate. Since the patient's subjective well-being is influenced by many factors unrelated to the disease process or treatment (i.e., education, income, quality of the environment, etc.), investigators have adopted the narrower term,

HRQL. Some definitions of HRQL stem from the recognition that HRQL may be considered on different levels: an overall assessment of well-being; several broad domains—physiological, functional, psychological, and social status; and subcomponents of each domain—for example pain, sleep, activities of daily living, sexual function within physical and functional domains.

It follows that HRQL is a multifactorial concept that, from the patient's perspective, represents the final common pathway of all the physiological, psychological, and social influences of the therapeutic process.⁴ It follows also that when assessing the impact of a drug on patients' HRQL, one may be interested in describing the patients' status (or changes in the patients' status) on a whole variety of domains, and that different strategies and instruments are required to explore separate domains.

Definitions of HRQL, both theoretical and practical, remain controversial. Most HRQL measurement instruments focus largely on how patients are functioning, e.g., their ability to care for themselves and carry out their usual roles in life. While this pragmatic view of HRQL has gained ascendancy, there remain those who argue that unless you are tapping into individual patients' values you may be measuring health status, but you are not measuring HRQL.⁵

These issues can be clarified by thinking of a woman with quadriplegia who, despite her limitations, is very happy and fulfilled and values her life highly (more, for instance, than most people, or than she did before she suffered quadriplegia). On most domains of most HRQL instruments, this woman's results would suggest a poor HRQL, despite the high value she places on her health state. Investigators and those interpreting the results of HRQL measure should be aware of the varying emphasis put on individual patient values in the different types of instrument.⁶

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

HRQL effects may be pertinent in both investigating and documenting 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 are to take the medication and live with both its beneficial actions and side effects. Investigators must therefore recognize the clinical situations where a drug may have an important effect on HRQL. This requires careful examination of data available from earlier phases of drug testing and, until now, has usually been performed in the latter stages of phase III testing. For example, Croog and colleagues studied the effect of three established antihypertensive drugs—captopril, methyl-dopa, and propranolol—on quality of life, long after their introduction in clinical practice.⁷ Their report, which showed an advantage of captopril in several HRQL domains, had a major impact on the drug prescription pattern at the time of its publication. The earlier in the process of drug development potential effects on quality of life are recognized, the sooner appropriate data may be collected and analyzed.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

Researchers willing to accept the notion of the importance of measuring HRQL in pharmacoepidemiology research and ready to use HRQL instruments in postmarketing (or, in some cases, premarketing) trials, face a considerable number of challenges. These challenges start with the realization that, as we have noted, there is no universal agreement on what the concept of quality of life actually entails. Thus, investigators must define as precisely as possible the aspects of HRQL in which they are interested.

Having identified the purpose for which a HRQL instrument is to be used, one must be aware of the measurement properties required for it to fulfill its purpose. An additional problem occurs at this stage if the original instrument was developed in a different language—the adequate performance of an instrument cannot be assumed after its translation. At the next step, the investi-

gator is challenged by the task of choosing from many available HRQL measurement instruments. When all these problems are dealt with satisfactorily, the investigator has to ensure that the measurements (interviews or self- or computer administered questionnaires) are made in a rigorous (standardized, reproducible, unbiased) fashion. 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 general terms, 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.^{8,9} In most clinical trials, quality of life measurement instruments are used for evaluation of the effects of therapy, with treatment effect being expressed as a change in the score of the instrument over time. Occasionally, instruments are used 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 an instrument is used 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 supposed 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 stepwise process including examination of *face validity* (or sensibility)¹⁰ 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. For example, one could hypothesize that changes in spirometry related to a use of a new drug in patients with chronic airflow limitation should bear a close correlation with changes in functional status of the patient and a weaker correlation with changes in their emotional status.

The second attribute of an HRQL instrument is its ability to detect the “signal,” over and above the “noise” is introduced in the measurement process. For *discriminative instruments*, those that measure differences among people at a single point in time, this “signal” comes from differences between patients in HRQL. In this context, the way of quantitating the signal-to-noise ratio is called *reliability*. If the variability in scores between 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.

An example of an index of responsiveness is the ratio of the magnitude of change that corresponds to the minimally important difference, to the variability in score in stable subjects. Alternatively, the minimally important difference can be related to the variability associated with measuring differences in subjects who are changing. Investigators have suggested other measurements of responsiveness, but they all rely on some way of relating signal to noise.^{11–14}

Another essential measurement property of an instrument is the extent to which one can understand the magnitude of any differences between treatments that a study demonstrates—the instrument's *interpretability*. If a treatment improves HRQL score by three points relative to control, what are we to conclude? Is the treatment effect very large, warranting widespread dissemination, or is it trivial, suggesting the new treatment should be abandoned? This question highlights the importance of being able to interpret the results of our HRQL questionnaire scores.

While our capacity to interpret results remains limited, investigators are adducing more and more information to enhance instrument interpretability. Researchers have developed a number of strategies to address this difficult issue. Successful strategies have three things in common. First, they require an independent standard of comparison. Second, this independent standard must itself be interpretable. Third, there must be at least a moderate relationship between changes in questionnaire score and changes in the independent standard. We have found that a correlation of 0.5 approximates the boundary between an acceptable and unacceptable relationship for establishing interpretability.

In our own work, we have often used global ratings of change (patients classifying themselves as unchanged, or experiencing small, medium, and large improvements or deteriorations) as the independent standard. We construct our disease-specific instruments using seven-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 one point change in two or more items as important. This finding seems to apply across different areas of function, including dyspnea, fatigue, and emotional function in patients with chronic airflow limitation;¹⁵ symptoms, emotional function, and activity limitations in both adult¹⁶ and child¹⁷ asthma patients, and parents of child asthma patients;¹⁸ and symptoms, emotional function, and activity limitations in adults with rhinoconjunctivitis.¹⁹ Similar observations may be derived from reports of others.²⁰

The approach that we have just described relies on within-patient comparisons as the independent standard. One 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. The patients conversed with one another long enough to make judgements 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 allows estimates of what differences in Chronic Respiratory Questionnaire score constitute small, medium, and large differences. The results were largely congruent with the findings from the within-patient rating studies.²¹

Investigators can also use measures that physicians, through long experience, already know well, as independent standards. For example, scores on a generic measure of HRQL, the Sickness Impact Profile (SIP), range from an average of 8.2 in patients with American Rheumatism Association arthritis class I, to 25.8 in class IV.²² Another standard would be obtained by administering questionnaires to patients before and after an intervention of known effectiveness with which clinicians are familiar, so that they can see the change in score associated with response to treatment. For example, patients shortly after hip replacement have scores of 30 on the SIP, scores which decrease to less than 5 after full convalescence.²³ Relationships between HRQL and a variety of marker states can also be useful: SIP scores in patients with chronic airflow limitation severe enough to require home oxygen are approximately 24;²⁴ scores in patients with chronic, stable angina are approximately 11.5.²⁵

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 a minimally important difference (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 had they received the control and ignores possible heterogeneity of treatment effect. Depending on the true distribution of results, the appropriate interpretation may be different.

Consider a situation where 25% of the treated patients improved by a magnitude of 1.0, while the other 75% did not improve at all (mean change of 0). This would mean that the 25% of treated patients obtained moderate benefit from the intervention. Using the methodology that has recently been developed for interpreting the magnitude of treatment effects, the number needed to treat (NNT), investigators have found that clinicians commonly treat 25 to 50 patients, and

often as many as 100, to prevent a single adverse event.^{26,27} Thus, the hypothetical treatment with a mean difference of 0.25 and an NNT of four proves to have a powerful effect.

We have shown that this issue is much more than hypothetical.²⁸ In a crossover randomized trial in asthmatic patients comparing the short acting inhaled β -agonist salbutamol to the long acting beta 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 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 2.2. The NNT is simply the inverse of the proportion who benefit, in this case 4.5. Thus, clinicians thus need to treat fewer than five patients with salmeterol to ensure that one patient obtains an important improvement in their ability to undertake activities of daily living.

In another randomized trial examining the effect of a respiratory rehabilitation program in patients with chronic lung disease, we found a mean difference between rehabilitation patients and the community controls of 0.40 in the emotions domain of the Chronic Respiratory Questionnaire. This difference is appreciably less than the value of 0.5 that represents the minimal important difference in an individual patient. However, the data from the trial allow us to calculate the proportion of patients who were 0.5 points or more better in their emotional function while receiving rehabilitation than would have been the case had they been in the community control group. This turns out to be 0.30, which translates into an NNT of 3.3 patients.

This discussion emphasizes that to interpret 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

During the last decade, clinical journals have started to publish trials in which HRQL instruments are the primary outcome measures. With the expanding importance of HRQL in evaluating new therapeutic interventions, investigators (and readers) are faced with a large array of instruments. Researchers have proposed different ways of categorizing these instruments, according to the purpose of their use, into instruments designed for screening, providing health profiles, measuring preference, and making clinical decisions,²⁹ or into discriminative and evaluative instruments (as above).

We have also suggested a taxonomy based on the domains of HRQL which an instrument attempts to cover.³⁰ According to this taxonomy, an HRQL instrument may be categorized, in a broad sense, as generic or specific. *Generic instruments* cover (or at least aim to cover) the complete spectrum of function, disability, and distress of the patient, and are applicable to a variety of populations. Within the framework of generic instruments, health profiles and utility measures provide two distinct approaches to measurement of global quality of life. *Specific instruments* are focused on disease or treatment issues specifically relevant to the question at hand.

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 12 "categories," which can be aggregated into two dimensions and five independent categories, and also into a single overall score.³¹ The SIP has been used in studies of cardiac rehabilitation,³² total hip joint arthroplasty,³³ and treatment of back pain.³⁴ In addition to the SIP, there are a number of other health profiles available: the Nottingham Health Profile,³⁵ the Duke-UNC Health Profile,³⁶ and the McMaster Health Index Questionnaire.³⁷ Increasingly, a collection of related instruments from the Medical Outcomes Study³⁸ have become the most popular and widely used generic instruments. Particularly popular is one version that includes 36 items, the SF-36.³⁹⁻⁴¹

While each health profile attempts to measure all important aspects of HRQL, they may slice the HRQL pie quite differently. For example, the McMaster Health Index Questionnaire follows the World Health Organization approach and identifies three dimensions: physical, emotional, and social. The Sickness Impact Profile includes a physical dimension (with categories of ambulation, mobility, body care, and movement), a psychosocial dimension (with categories including social interaction and emotional behavior), and five independent categories including eating, work, home management, sleep and rest, and recreations and pastimes.

General health profiles offer a number of advantages to the clinical investigator. Their reproducibility and validity have been established, often in a variety of populations. When using them for discriminative purposes, one can examine and establish areas of dysfunction affecting a particular population. Identification of these areas of dysfunction may guide investigators who are constructing disease-specific instruments to target areas of potentially greatest impact on the quality of life. Health profiles, used as evaluative instruments, allow determination of the effects of an intervention on different aspects of quality of life, without necessitating the use of multiple instruments (and thus saving both the investigator's and the patient's time). Because health profiles are designed for a wide variety of conditions, one can

potentially compare the effects on HRQL of different interventions in different diseases. Profiles that provide a single score can be used in a cost-effectiveness analysis, in which the cost of an intervention in dollars is related to its outcome in natural units (see Chapter 35).

The main limitation of health profiles is that they may not focus adequately on the aspects of quality of life specifically influenced by a particular intervention. This may result in an inability of the instrument to detect a real effect in the area of importance (i.e., lack of responsiveness). We will return to this issue when we discuss the alternative approach, specific instruments.

Utility Measurement

Economic and decision theory provides the underlying basis for *utility measures* (see Chapter 35). The key elements of an 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 death (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. 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 (such as a random sample of the general population). This approach is typified by two widely used instruments, the Quality of Well-Being Scale⁴²⁻⁴⁴ and the Health Utilities Index.⁴⁵

The second approach is to ask patients to make a single rating which takes into account all aspects of their quality of life.⁴⁶ This rating can be made 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 major advantage of utility measurement is its amenability to cost-utility analysis (see Chapter 35). 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 provide a basis for allocation of scarce resources among different healthcare programs. Results from the utility approach may thus be of particular interest to program evaluators and health policy decision makers.

However, utility measurement also has limitations. Utilities can vary depending on how they are obtained, raising questions of the validity of any single measurement.^{47,48} Utility measurement does not allow the investigator to determine which aspects of HRQL are responsible for changes in utility. Finally, utilities potentially share the disadvantage of health profiles, in that they may not be responsive to small but still clinically important changes.

SPECIFIC INSTRUMENTS

An alternative approach to HRQL measurement is to focus on aspects of health status that are specific to the area of primary interest. The rationale for this approach lies in the increased responsiveness that may result from including only those aspects of HRQL that are relevant and important in a particular disease process or even in a particular patient situation. One could also focus an instrument only on the areas that are likely to be affected by a particular drug. This latter approach is advanced in the design and conduct of randomized controlled trials in individual patients—*N-of-1* randomized clinical trials⁴⁹ (see Chapter 37).

In other situations, the instrument may be specific to the disease (instruments for chronic lung disease, for rheumatoid arthritis, for cardiovascular diseases, for endocrine problems, etc.); specific to a population of patients (instruments

designed to measure the HRQL of the frail elderly, who are afflicted with a wide variety of different diseases); specific to a certain function (questionnaires that examine emotional or sexual function); or specific to a given condition or problem (such as pain) that can be caused by a variety of underlying pathologies. Within a single condition, the instrument may differ depending on the intervention. For example, while success of a disease modifying agent in rheumatoid arthritis should result in improved HRQL by enabling a patient to increase performance of physically stressful activities of daily living, occupational therapy may achieve improved HRQL by encouraging family members to take over activities formerly accomplished with difficulty by the patient. Appropriate disease-specific HRQL outcome measures should reflect this difference.

Specific instruments can be constructed to reflect the "single state" (how tired have you been: very tired, somewhat tired, full of energy) or a "transition" (how has your tiredness been: better, the same, worse).⁵⁰ Theoretically, the same could be said of generic instruments, although none of the available generic instruments has used the transition approach. Specific measures can integrate aspects of morbidity, including events such as recurrent myocardial infarction.⁵¹

The disease-specific instruments may be used for discriminative purposes. They may aid, for example, in evaluating the extent to which a primary symptom (for example dyspnea) is related to the magnitude of physiological abnormality (for example exercise capacity).⁵² Disease-specific instruments can be applied for evaluative purposes to establish the impact of an intervention on a specific area of dysfunction, and hence aid in elucidating the mechanisms of drug action.⁵³ Guidelines provide structured approaches for constructing specific measures.⁵⁴ 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.⁵⁵⁻⁶⁰

In addition to the likelihood of 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.⁶¹ Specific measures may therefore appear clinically sensible to the physician.

The disadvantages of specific measures are that they are (deliberately) not comprehensive, and cannot be used to compare across conditions or, at times, even across programs. This suggests that there is no one group of instruments that will achieve all the potential goals of HRQL measurement. Thus, investigators may choose to use multiple instruments, an issue we will deal with in the next section.

USE OF MULTIPLE QUALITY OF LIFE MEASURES IN CLINICAL STUDIES

Clinical investigators are not restricted to using a single instrument in their studies, and investigators will often conclude that a single instrument cannot yield all the relevant information. For example, utility and disease-specific measures contribute quite different sorts of data, and an investigator may want to use one of each.

Another, somewhat different way of using multiple instruments is to administer a battery of specific instruments. An example of such an approach was a double blind, randomized trial of three antihypertensive agents in primary hypertension.⁷ The investigators identified five dimensions of health they were measuring: the sense of well-being and satisfaction with life, the physical state, the emotional state, intellectual functioning, ability to perform in social roles, and the degree of satisfaction from those roles. Even within these five dimensions, additional components were identified. For example, separate measurements of sleep and sexual function were made. Patients taking one of the three drugs under investigation, captopril, scored better on measures of general well-being, work performance, and life satisfaction. The lesson for the clinician is clearly important: one can have an impact on not only the length, but also the quality of the patient's life according to choice of antihypertensive agent.

This approach, although comprehensive, has limitations. First, investigators must find a valid, responsive instrument for every attribute they wish to measure. Second, it is possible (indeed likely) that only some of the instruments chosen will show differences between the treatments under investigation. Unless one of the instruments has been designated as the primary measure of outcome before the study started, different results in different measures may make interpretation difficult. The greater the number of instruments used, the greater the probability that one or more will favor one treatment or the other, even if the treatments' true effectiveness is identical. Thus, the α error (the probability of finding an apparent difference between treatments when in fact their outcomes do not differ) increases with each new instrument used. Although this problem may be dealt with through statistical adjustment for the number of instruments used, such adjustment is often not made.⁶²

Another problem occurs if only a small proportion of the instruments used favor an intervention (or if some measures favor one treatment and other instruments favor the other). In these situations, the clinician may be unsure how to interpret the results. The use of multiple instruments opens the door to such potential controversy.

A final limitation of using a battery of instruments is that it gives no indication of the relative importance of various areas of dysfunction to the patient. For example, had Croog *et al.*⁷ found that one antihypertensive agent disturbed sleep, while another had an adverse impact on sexual function, their approach would not have allowed determination of which drug had a greater net adverse impact on patients' lives.

THE FUTURE

The considerations we have raised suggest a step-by-step approach to addressing issues of HRQL in pharmacoepidemiology studies.⁶³ Clinicians must begin by asking themselves whether investigators have addressed all the important effects of treatment on patients' quantity and quality of life.^{64,65}

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 right instruments? In particular, does evidence suggest the measure(s) used are valid measures of HRQL? If so, and the study failed to demonstrate differences between groups, is there good reason to believe the instrument is responsive in this context? If not, the results may be a false negative, failing to show the true underlying difference in HRQL.

Whatever the differences between groups, the clinician must be able to interpret their magnitude. Knowledge of the difference in score that represents small, medium, and large differences in HRQL will be very helpful in making this interpretation. Clinicians must still look beyond mean differences between groups, and consider the distribution of differences. The number of patients needed to treat to ensure that a single patient achieves an important benefit in HRQL offers one way of expressing results that clinicians are likely to find meaningful.

REFERENCES

1. Franciosa JA, Jordan RA, Wilen MM *et al.* Minoxidil in patients with chronic left heart failure: contrasting hemodynamic and clinical effects in a controlled trial. *Circulation* 1984; **70**: 63–8.
2. Packer M. How should we judge the efficacy of drug therapy in patients with chronic congestive heart failure? The insight of the six blind men. *J Am Coll Cardiol* 1987; **9**: 433–8.
3. Danis M, Gerrity MS, Southerland LI, Patrick DL. A comparison of patient, family, and physician assessment of the value of medical intensive care. *Crit Care Med* 1988; **16**: 594–600.
4. Schipper H, Clinch J, Powell V. Definitions and conceptual issues. In: Spilker B, ed., *Quality-of-Life Assessment in Clinical Trials*. New York: Raven, 1990; 11–24.
5. Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *J Am Med Assoc*, 1994; **272**: 619–26.
6. Guyatt GH, Cook DJ. Health Status, Quality of Life, and the Individual Patient. A Commentary on: Gill TM, Feinstein AR. A Critical Appraisal of Quality of Life Measurements. *J Am Med Assoc* 1994; **272**: 630–1.

7. Croog SH, Levine S, Testa MA *et al*. The effects of antihypertensive therapy on the quality-of-life. *New Engl J Med* 1986; **314**: 1657–64.
8. Kirshner B, Guyatt GH. A methodologic framework for assessing health care indices. *J Chron Dis* 1985; **38**: 27–36.
9. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992; **45**: 1341–5.
10. Feinstein AR. The theory and evaluation of sensibility. In: Feinstein AR, ed., *Clinometrics*. New Haven: Yale University Press, 1987; 141–65.
11. Levine MN. Incorporation of quality-of-life assessment into clinical trials. In: Osoba D, ed., *Effect of Cancer on Quality-of-Life*. Boston, MA: Chemical Rubber Company, 1991; 105–11.
12. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life: Basic Sciences Review. *Ann Intern Med* 1993; **70**: 225–30.
13. Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987; **40**: 171–8.
14. Osoba D, Aaronson NK, Till JE. A practical guide for selecting quality-of-life measures in clinical trials and practice. In: Osoba D, ed., *Effect of Cancer on Quality-of-Life*. Boston, MA: Chemical Rubber Company, 1991; 89–104.
15. 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.
16. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994; **47**: 81–7.
17. Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in children with Asthma. *Qual Life Res*. 1996; **5**: 35–46.
18. Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in the parents of children with asthma. *Qual Life Res* 1996; **5**: 27–34.
19. Juniper EF, Guyatt GH, Griffith LE, Ferrie PJ. Interpretation of rhinoconjunctivitis quality of life questionnaire data. *J Allergy Clin Immunol* 1996; **98**: 843–5.
20. Laupacis A, Wong C, Churchill D, and the Canadian Erythropoietin Study Group. The use of generic and specific quality-of-life measures in hemodialysis patients treated with Erythropoietin. *Control Clin Trials* 1991; **12** (suppl): 168S–179S.
21. Redelmeier DA, Goldstein RS, Guyatt GH. Assessing the Minimal Important Difference in Symptoms: A Comparison of Two Techniques. *J Clin Epidemiol* 1996; **49**: 1215–9.
22. Deyo RA, Inui TS, Leininger JD, Overman SS. Measuring functional outcomes in chronic disease: a comparison of traditional scales and a self-administered health status questionnaire in patients with rheumatoid arthritis. *Med Care* 1983; **21**: 180–92.
23. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: Development and Final Revision of a Health Status Measure. *Med Care* 1981; **19**: 787–805.
24. McSweeney AJ, Grant I, Heaton RK *et al*. Life quality of patients with chronic obstructive pulmonary disease. *Arch Intern Med* 1982; **142**: 473–8.
25. Fletcher A, McLoone P, Bulpitt C. Quality of life on angina therapy: a randomised controlled trial of transdermal glyceryl trinitrate against placebo. *Lancet* 1988; **2**: 4–7.
26. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *New Engl J Med* 1988; **318**: 1728–33.
27. Jaeschke R, Guyatt G, Shannon H, Walter SD, Cook DJ, Hedde N. Basic Statistics for Clinicians. iii. Assessing the Effects of Treatment: Measures of Association. *Can Med Assoc J* 1995; **152**: 351–7.
28. Guyatt GH, Juniper EL, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomized trials. *Br Med J*. 1998; **316**: 690–3.
29. Osoba D. *Effect of Cancer on Quality-of-Life*. Boston, MA: Chemical Rubber Company, 1991.
30. Guyatt GH, Veldhuyzen Van Zanten SJO, Feeny DH, Patrick DL. Measuring quality of life in clinical trials: a taxonomy and review. *Can Med Assoc J* 1989; **140**: 1441–7.
31. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981; **19**: 787–805.
32. Ott CR, Sivarajan ES, Newton KM *et al*. A controlled randomized study of early cardiac rehabilitation: the Sickness Impact Profile as an assessment tool. *Heart Lung* 1983; **12**: 162–70.
33. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985; **28**: 542–7.
34. Deyo RA, Diehl AK, Rosenthal M. How many days of bed rest for acute low back pain? a randomized clinical trial. *New Engl J Med* 1986; **315**: 1064–70.
35. Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health* 1980; **34**: 281–6.
36. Parkerson GR, Gehlback SH, Wagner EH, Sherman AJ, Clapp NE, Muhlbaier LH. The Duke–UNC Health Profile: an adult health status instrument for primary care. *Med Care* 1981; **19**: 806–28.

37. Sackett DL, Chambers LW, MacPherson AS, Goldsmith CH, McAuley RG. The development and application of indices of health: general methods and a summary of results. *Am J Public Health* 1977; **67**: 423–8.
38. Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study. *J Am Med Assoc* 1989; **262**: 925–30.
39. Brook RH, Ware JE, Davies-Avery A, Stewart AL, Donald CA *et al*. Overview of adult health status measures fielded in Rand's health insurance study. *Med Care* 1979; **17** (suppl 7): iii–55.
40. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). *Med Care* 1992; **30**: 473–83.
41. Ware JE, Kosinski M, Bayliss MS *et al*. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: Summary of results of the Medical Outcomes Study. *Med Care* 1995; **33**: AS264–AS279.
42. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. *Health Serv Res* 1976; **11**: 478–507.
43. Kaplan RM, Bush JW. Health-related quality-of-life measurement for evaluation research and policy analysis. *Health Psychol* 1982; **1**: 61–80.
44. Kaplan RM, Anderson JP, Wu A, Mathews WC, Kozin F *et al*. The Quality of Well-Being Scale: applications in AIDS, cystic fibrosis, and arthritis. *Med Care* 1989; **27** (suppl): S27–S43.
45. Feeny DH, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems: Health Utilities Index. *Pharmacoeconomics* 1995; **7**: 490–502.
46. Torrance GW. Measurement of health state utilities for economic appraisal: a review. *J Health Econ* 1986; **5**: 1–30.
47. Sutherland HJ, Dunn V, Boyd NF. Measurement of values for states of health with linear analog scales. *Med Decis Making* 1983; **3**: 477–87.
48. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R *et al*. The measurement of patients' values in medicine. *Med Decis Making* 1982; **2**: 449–62.
49. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The N-of-1 randomized controlled trials: clinical usefulness. Three year experience. *Ann Intern Med* 1990; **112**: 293–9.
50. MacKenzie CR, Charlson ME. Standards for the use of ordinal scales in clinical trials. *Br Med J* 1986; **292**: 40–3.
51. Olsson G, Lubsen J, Van Es G, Rehnqvist N. Quality-of-life after myocardial infarction: effect of long term metoprolol on mortality and morbidity. *Br Med J* 1986; **292**: 1491–3.
52. Mahler DA, Rosiello RA, Harver A, Lentine T, McGovern JF, Daubenspeck JA. Comparison of clinical dyspnea ratings and psychophysical measurements of respiratory sensation in obstructive airway disease. *Am Rev Respir Dis* 1987; **135**: 1229–33.
53. Jaeschke R, Singer J, Guyatt GH. Using quality-of-life measures to elucidate mechanism of action. *Can Med Assoc J* 1991; **144**: 35–9.
54. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality-of-life in clinical trials. *Can Med Assoc J* 1986; **134**: 889–95.
55. Tandon PK, Stander H, Schwarz RP Jr. Analysis of Quality of Life data from a Randomized, Placebo controlled Heart-Failure trial. *J Clin Epidemiol* 1989; **42**: 955–62.
56. Smith D, Baker G, Davies G, Dewey M, Chadwick DW. Outcomes of add-on treatment with Lamotrigine in partial epilepsy. *Epilepsia* 1993; **34**: 312–22.
57. Chang SW, Fine R, Siegel D, Chesney M, Black D, Hulley SB. The Impact of Diuretic Therapy on Reported Sexual Function. *Arch Intern Med* 1991; **151**: 2402–8.
58. Tugwell P, Bombardier C, Buchanan WW, Goldsmith C, Grace E, Bennett KJ *et al*. Methotrexate in Rheumatoid Arthritis. Impact on quality of life assessed by traditional standard-item and individualized patient preference health status questionnaires. *Arch Intern Med* 1990; **150**: 59–62.
59. Laupacis A, Wong C, Churchill D. The use of generic and specific quality-of-life measures in hemodialysis patients treated with erythropoietin. *Control Clin Trials* 1991; **12**: 168S–179S.
60. Goldstein RS, Gort EH, Guyatt GH, Stubbing D, Avendano MA. Prospective randomized controlled trial of respiratory rehabilitation. *Lancet* 1994; **344**: 1394–7.
61. 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.
62. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. *New Engl J Med* 1987; **317**: 426–32.
63. Guyatt GH, Naylor CD, Juniper EL, Heyland DK, Jaeschke RZ, Cook DJ. Users' Guides to the Medical Literature: XII. How to use articles about health-related quality of life. *J Am Med Assoc* 1997; **277**: 1232–7.
64. Guyatt GH, Sackett DL, Cook DJ for the Evidence-based Medicine Working Group. Users' guides to the medical literature: II. How to use an article about therapy or prevention: Part B. What were the results and will they help me in caring for my patients. *J Am Med Assoc* 1994; **271**: 59–63.
65. Dans AL, Dans LF, Guyatt GH, Richardson S, for the Evidence-based Medicine Working Group. User's guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. *J Am Med Assoc*. 1998; **279**: 545–549.

37

N-of-1 Randomized Clinical Trials in Pharmacoepidemiology

GORDON H. GUYATT

Department of Medicine and Department of Clinical Epidemiology and Biostatistics, McMaster University,
Hamilton, Ontario, Canada

ROMAN JAESCHKE

Department of Medicine, McMaster University, Hamilton, Ontario, and Department of Medicine, St. Joseph's
Hospital, Hamilton, Ontario, Canada

ROBIN ROBERTS

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

INTRODUCTION

Clinicians are used to, and comfortable with, making initial clinical recommendations for their patients on the basis of medical research consisting of formal randomized clinical trials of groups of patients (see Chapter 2). However, individual patients may not behave the same way as the group. Responding to this heterogeneity of treatment effect, psychological research has long taken advantage of experimental studies of single subjects. In this chapter, we describe how clinicians can use an experimental approach that focuses on the individual to facilitate clinical research and, especially, to enhance the clinical care of their patients. We believe the approaches we describe are highly relevant to pharmacoepidemiology

because they apply fundamental epidemiological principles to the selection of drugs for individual patients, and to the drug development process.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

Traditionally, important questions that patients, and clinicians caring for them, consider include: does treatment have the potential to eradicate the disease?; what is the probability of prolonging the life of the patient?; what is the chance of delaying the progression of a disease?; what is the frequency and impact of a drug's side effects? Randomized

controlled trials (RCTs) address these questions, and are usually necessary to establish valid evidence of drug efficacy¹ (see also Chapters 1, 2, and 34). When deciding which therapy is more beneficial for an individual patient, however, clinicians often cannot rely on the results of RCTs. For example, an RCT addressing the issue may not be available; some conditions are so rare that even multicenter collaborative trials are not feasible. Further, even when a relevant RCT generated a clear answer, its result may not apply to an individual patient. First, if the patient does not meet the eligibility criteria, extrapolation may not be appropriate.² Second, regardless of the overall trial results, some patients may benefit from a given therapy while others do not.

Under these circumstances, clinicians typically conduct the time-honored “trial of therapy,” in which the patient is given a treatment and the subsequent clinical course determines whether the treatment is judged effective and continued. However, many factors may mislead physicians conducting conventional therapeutic trials. They include the placebo effect, the natural history of the illness, the expectations that the clinician and patient have about the treatment effect, and the desire of the patient and the clinician not to disappoint one another. To avoid these pitfalls, clinicians must conduct trials of therapy with safeguards that keep both patients and themselves “blind” to the treatment being administered. Investigators routinely use such safeguards in large scale RCTs involving dozens or hundreds of patients. To implement these safeguards in a clinically sensible manner while investigating drug effects in the individual patient constitutes a considerable challenge.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

To maintain the methodological safeguards provided by RCTs and to avoid the disadvantages of large sample multicenter studies, we built on the work of experimental psychologists to develop a

corresponding methodology for examination of an intervention's effect in individual patients in clinical practice.³ Any person conducting an RCT in an individual patient (*N*-of-1 RCT) faces several formidable methodological challenges. The first one is to identify a clinical situation in which an *N*-of-1 RCT may potentially provide useful information. The second is to properly assess the patient's (and clinician's) willingness to undergo the time and effort consuming process of the *N*-of-1 RCT. The third is to choose appropriate treatment targets to monitor the effects of therapy (both beneficial and potentially harmful). Once these elements are in place, the clinician–patient team has to design, execute, and analyze the results of the trial. The clinician must carefully think through each of these elements before commencing an *N*-of-1 RCT.

CURRENTLY AVAILABLE SOLUTIONS

N-OF-1 RCT: THE GENERAL CASE

The methodology of experimental studies of single patients is known as “single case” or “single subject” research, $N = 1$, or, as we call it, *N*-of-1 RCTs.^{4,5} We have previously described how *N*-of-1 RCTs may be used in medical practice to determine the optimum treatment of an individual patient, described an “*N*-of-1 service” designed to assist clinicians who wish to conduct such a trial, provided detailed guidelines for clinicians interested in conducting their own *N*-of-1 RCTs, and reviewed our own three years' experience in conducting such studies.^{6,7} In each of two conditions (chronic airflow limitation and fibromyalgia) we conducted over 20 *N*-of-1 RCTs, and described our experience with these patients in two separate reports.^{8,9}

In general terms, the *N*-of-1 RCT design is based on pairs of active/placebo, high dose/low dose, or first drug/alternate drug combinations, the order of administration within each pair determined by random allocation (see Figure 37.1). Treatment targets (directed specifically at the patient's complaints) are monitored in a double blind fashion on a regular, predetermined

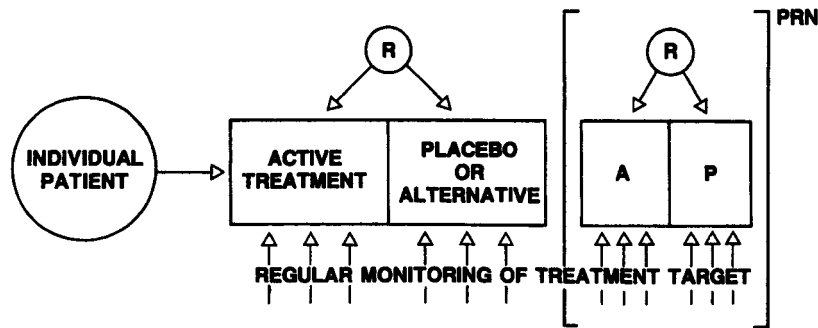


Figure 37.1. *N*-of-1 study design.

schedule. The trial continues as long as the clinician and patient agree that they need more information to obtain a definite answer regarding the efficacy, superiority, or side-effects of the treatment, or until the patient or clinician decide for any other reason to end the trial. Experience drawn from conducting numerous *N*-of-1 RCTs indicates that one should usually aim for at least three treatment pairs. If a very short duration of treatment periods is feasible, the number of pairs could be greater.

Several criteria must be satisfied before clinicians begin an *N*-of-1 RCT. In addition to the effectiveness or harmful side-effects of treatment being in doubt, the disorder should be relatively chronic and stable. The treatment, if effective, should be continued relatively long term, and the patient should be eager to collaborate in designing and carrying out the *N*-of-1 RCT. The treatment(s) must have a rapid onset and termination of action, and the clinician should be aware of the optimal treatment duration, which must be practical. Last but not least, the clinician must secure the cooperation of a pharmacy in the preparation of blinded medication.

N-OF-1 RCT: CHOOSING APPROPRIATE OUTCOME MEASURES

The choice of the target to be monitored during *N*-of-1 RCTs depends on the clinical context. Although one may choose physiological variables as treatment targets (say forced expired volume in one second in an RCT of a bronchodilator, or postural blood pressure change in a trial of

amitriptyline in symptomatic postural hypotension), *N*-of-1 RCTs are uniquely suitable in sorting out the effects of a given drug on different aspects of health related quality of life (HRQL, see Chapter 36). Let us assume that the problem with which the clinician is faced is new onset of pain associated with walking (intermittent claudication) in a patient with peripheral vascular disease and a previous myocardial infarction. This patient began taking β -blockers some time ago to prevent a recurrent myocardial infarction. Now the clinician reads the insert to the β -blocker and confirms what she remembers from medical school: β -blockers may exacerbate the symptoms of peripheral vascular disease. She then finds a meta-analysis that concludes that β -blockers do not have this effect.¹⁰ A clinical dilemma ensues—should the patient (free of symptoms of claudication before being put on β -blocker) stay on the medication or should it be discontinued? This situation fulfills all the requirements necessary for conduct of an *N*-of-1 RCT: the condition is chronic; the potential effect of the β -blocker on peripheral vasculature should have rapid onset and termination of action; and the drug, if proven not to cause the troublesome symptoms, will be continued long term.

The clinician should establish the treatment targets most important for the patient before the trial, tailoring the choice to the individual case. The patient may be concerned about his ability to walk a dog, visiting a family who lives on the fourth floor of a building with no elevator, the severity of his discomfort when playing golf, or embarrassment due to other people noticing problems with walking. The clinician must moni-

tor whatever symptoms the patient judges as sufficiently important and troublesome that getting rid them may be worth a 25% increase in the risk of myocardial infarction.

The simplest outcome measure could be the preference for β -blocker or placebo in a pair of treatment periods. If one would like to quantify the degree of symptoms and explore the reasons behind the patient's preference, the response options depend on the nature of the treatment target. We have found seven-point scales an optimal method for recording the severity of symptoms because of their ability to detect small degrees of change, patients' ease in understanding them, and their ease of interpretation.¹¹ Using a seven-point scale response, for example, one may ask

Please indicate how much leg discomfort have you been experiencing while playing golf last week by choosing one of the options from the scale below:

- 1 Extreme discomfort
- 2 Severe discomfort
- 3 Quite a lot of discomfort
- 4 Moderate discomfort
- 5 Mild discomfort
- 6 A little discomfort
- 7 No discomfort at all.

Using the taxonomy introduced in the chapter on quality of life studies, we call a set of such questions a disease- and patient-specific questionnaire (see Chapter 36). Other options for monitoring treatment targets include other disease-specific or general quality of life questionnaires.

To summarize our hypothetical example, one could conduct a *N*-of-1 RCT consisting of three treatment pairs, with each treatment pair including a two week treatment period on β -blocker and a two week period on placebo. The patient could complete the outcome measures (a disease- and patient-specific three to five item symptom questionnaire) twice during the second week of each treatment period (second week only as the results during the first week could be influenced by the medication taken during the preceding treatment period).

N-OF-1 RCT: STATISTICAL INTERPRETATION OF THE RESULTS

An Example Trial

To illustrate issues in the interpretation of data from *N*-of-1 trials, we will use an example of an *N*-of-1 RCT in which we tested the effectiveness of ipratropium bromide in a dose of four inhalations four times a day (each inhalation 20 μ g of the drug) in a patient with chronic airflow limitation. Each day, the patient rated the severity of a number of symptoms, judged by him as important in his daily life. The symptoms included shortness of breath while walking upstairs, playing the vibraphone, drying the dishes, and overall shortness of breath. Each symptom was rated on a seven point scale in which a higher score represented better function (i.e., "1", extremely short of breath; ...; "7", not at all short of breath). The treatment periods were seven days in length (with measurements obtained during last five days of each period), and the trial included four pairs of treatment periods. Table 37.1 presents the mean scores for each of the eight weeks of the study.

Interpreting the Results: Visual Inspection

One way of evaluating these data is visual inspection. Visual inspection of the data is based on a graphical display of outcome assessments over time and in relation to the treatment being received. A conclusion that the treatment is effective would be supported to the extent that there is minimal variability within periods, that the difference between active and placebo periods is large relative to the within-period variability that is seen, and to the extent that the magnitude of the difference between active and placebo periods is consistent. Using the visual inspection of data one tries to intuitively draw conclusions about direction, magnitude, and consistency of response to experimental treatment. We invite readers to make conclusions on the basis of the data presented in Table 37.1 and Figure 37.2 before considering a more formal analysis.

Visual inspection is appealing in that it makes intuitive sense to both patient and clinician. Yet,

Table 37.1. Results of an N-of-1 RCT: ipratropium, symptom score (data represent mean score on four questions)

Pair	Treatment	Day 1	Day 2	Day 3	Day 4	Day 5	Treatment mean
1	Placebo	3.25	3.25	3.00	3.00	3.00	3.10
	Active	5.00	5.25	4.75	5.25	5.25	5.10
	Difference						+2.00
2	Placebo	3.75	3.25	3.25	2.75	2.75	3.15
	Active	4.50	4.50	4.75	4.50	4.70	4.59
	Difference						+1.44
3	Placebo	2.75	2.75	3.00	2.25	3.25	2.80
	Active	2.50	3.25	3.25	2.75	3.00	3.15
	Difference						+0.35
4	Placebo	3.00	3.00	2.75	2.75	2.75	2.85
	Active	4.50	4.50	5.00	5.25	—	4.81
	Difference						+1.96
All Pairs	Placebo						2.975
	Active						4.413
	Difference						+1.438

$t_3 = 1.438 / (0.7686 / 4) = 3.74$
[NB $t_3 = 14.00$]

visual inspection is subjective and thus may allow inconsistency in the evaluation of intervention effects. Use of the criteria presented above will help reduce subjectivity and thus increase the likelihood of consistent interpretation. However, even experts in visual analysis often disagree about particular data patterns and whether the effects are reliable.¹²

Interpreting the Results: Nonparametric Tests

An alternative approach to analysis of data from N-of-1 RCTs is to utilize a statistical test of significance, and a number of tests are possible candidates. These tests fall into two main classes: nonparametric and parametric tests. The major problematic issue in the interpretation of the

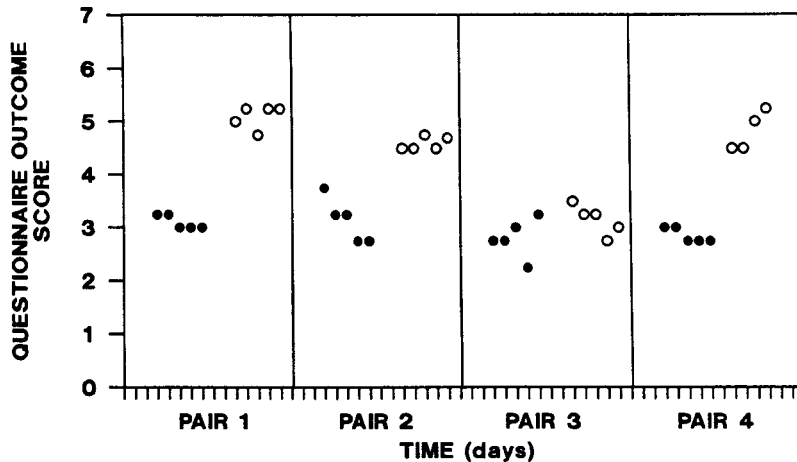


Figure 37.2. Results of an N-of-1 RCT: a visual display of mean daily symptom scores in a trial of ipratropium bromide versus placebo.

results of *N*-of-1 RCTs is that the errors associated with each data point may be correlated with the errors associated with other data points. This is called “autocorrelation” or “serial dependence” and we will address the issue in detail later in the discussion.

Like most statistical comparisons, nonparametric tests start from the hypothesis that treatment has no effect. An apparent effect of treatment observed in an *N*-of-1 RCT may well be due to a quirk of the order in which treatments were allocated within pairs. To evaluate this possibility, one considers the apparent treatment effects that would result from all other randomized treatment orders that could have occurred. The proportion of randomizations that produce apparent treatment effects as or more extreme than that actually observed is the *p*-value of the test.

In its simplest form, a randomization test would be based on the direction (i.e., sign) of the observed treatment difference in each pair. One can compute the *p*-value for the sign test which is appropriate for this form of randomization test from the binomial distribution or obtained from published tables.¹³ Using this approach for the analysis of the results of the ipratropium study described above, the first step is to calculate the mean score for each treatment period (presented in the last column of Table 37.1). In each pair of periods, the mean score favored the active treatment. Using the binomial theorem, the probability of this result occurring by chance if the treatment was ineffective is $1/2 \times 1/2 \times 1/2 \times 1/2 = 1/16$ (or 0.0625 for a one-sided test).

Two other nonparametric tests that incorporate the size of the outcome score differences are also commonly used. The Wilcoxon signed rank test¹⁴ uses the ranks of score differences. The absolute differences within pairs are first ordered from smallest to largest and ranks given to the difference, then if the two treatments are equivalent we would expect that the sum of ranks for positive differences should equal sum of negative ranks.

One can utilize a third option, a “pure” quantitative randomization test,¹⁵ by building up the empirical distribution of the difference in mean treatment outcome over all possible randomiza-

tions of the observed data. Under the null hypothesis of no systematic difference, the distribution of treatment mean difference would be centered on zero, but particular random treatment orderings would have led to variability in mean difference in both the positive and negative directions. The *p*-value of the test is again simply the proportion of randomizations that would have led to treatment differences as or more extreme than the one actually observed in the trial itself. This randomization test takes into account not only the direction of the difference (as the sign test does) and its relative size (as signed rank test does), but also its numerical size.

If one uses randomization tests, to obtain conventional one-tailed statistical significance one must complete at least five pairs and hope to obtain the most extreme result, because $1/2^5 = 1/32 = 0.0313$, while $1/2^4 = 1/16 = 0.0625$. The advantage of tests that quantify the size of the paired treatment differences rather than just their sign is manifest when one does not observe this extreme result. Whereas the sign test immediately drops to a *p* of 6/32 (0.1875) if only four of five pairs favor active treatment, the signed rank and, particularly, the pure randomization test can yield lower *p*-values (perhaps as low as 2/32) if the one pair favoring placebo does so only marginally. Thus, the expected *p*-values associated with an efficacious treatment tend to decrease from the sign test through the signed rank test to the randomization test, as each one is using more information. Tables for the Wilcoxon signed rank test are available,¹⁴ but the randomization test approach depends upon the actual data observed and must be calculated anew (either by hand or with a computer program) for each application.

Interpreting the Data: Parametric Tests

While nonparametric tests are very appealing in the context of an *N*-of-1 trial, their avoidance of strong distributional assumptions about the data make them relatively conservative compared to tests which make such assumptions. More powerful techniques for normally distributed outcomes, which we will refer to generally as analysis of variance, include analysis of variance itself and, as

a special case, the Student *t*-test (either in the paired or unpaired form).

In the ipratropium trial, there were four pairs of treatment periods and five symptom scores per period. If one assumes that the day-to-day fluctuation in outcome within treatment period is simply inherent random variation in the disease process one would call this, in the parlance of analysis of variance, a balanced two factor crossed design with replication. The factor of interest is treatment, but we must also allow for possible shifts in mean from one pair of periods to another and thus period pair represents the second factor. Less obvious to the statistical lay person is the need to allow for the possibility of interaction between treatment and period pair. In this situation, the interaction measures the tendency of the treatment effect to vary in size from one period pair to another.

The results of the analysis of variance with the imputed missing values are given in Table 37.2. The mean square for between pairs measures the variation in mean symptom score between period pairs, the mean square for between treatments reflects consistent differences in mean between active and placebo periods, and the mean square for interaction reflects the variation in active–placebo mean difference over period pairs. It is crucial to realize that period pair is a random factor in this design; in other words, that the four pairs of periods in which the study was run can be

thought of as a sample from a much larger population of time periods that could have been used. Because of this, the appropriate test of treatment compares the average difference between active and placebo means to the variability in active–placebo differences over treatment pairs, in the ratio of the mean square treatment to mean square interaction. The resulting *F* test ($F^{1,3} = 13.99, p = 0.033$) is identical to the square of the paired Student *t*-test conducted on the four pairs of treatment means.

One might reasonably question why one should use ANOVA here when the result is identical to the simple and more familiar paired *t*-test. The answer lies in the second test provided by the ANOVA, namely the test of interaction, which is computed as the ratio of the mean square interaction to the mean square error ($F(3, 31) = 19.03, p < 0.0001$). This formally asks the question, “Is there evidence of extra variation in treatment effect from period pair to period pair over and above that we would expect from the inherent day-to-day variation in symptom score?” If evidence for this extra variation were very weak, we would gain great advantage statistically by comparing the mean square treatments to the mean square error (or even the pooled error incorporating the interaction). Although the numerical size of *F* would not change much, the gain in degrees of freedom could have an appreciable effect on the *p*-value. In essence, the switch from using the yardstick of the

Table 37.2. Analysis of variance: ipratropium trial symptom scores

Source	df	SS	MS	<i>F</i>
Between pairs	3	7.32	2.44	$\frac{2.44}{0.88} = 31.33$ (<i>p</i> < 0.0001)
Between treatments	1	20.68	20.68	$\frac{20.68}{1.48} = 13.99$ (<i>p</i> = 0.033)
Interaction treatment × pairs	3	4.44	1.48	$\frac{1.48}{0.08} = 19.03$ (<i>p</i> < 0.0001)
Error	31	2.41	0.08	
Total	38	34.84		
Paired <i>t</i> approach:		Mean active–placebo difference = 1.438 SD (active–placebo difference) = 0.7686		

interaction mean square to the error mean square in testing the treatment effect allows us to use the variation in all 40 individual observations, rather than only the treatment mean differences in the four pairs of periods.

The ipratropium trial shows relatively strong evidence of extra variation through the test of interaction ($F(3, 31) = 19.03, p < 0.0001$). Thus, in this particular case, one could have used the simple paired *t*-test. However, we have used this approach in 16 *N*-of-1 trials to look at serial dependence. We found clear evidence of extra variation in treatment effect over period pairs in about 50% of trials. The additional variation could be a reflection of the long term ebb and flow of the disease. Alternatively, isolated, short term, truly random changes in symptom severity could also be responsible. In any case, the presence of the interaction dictates that, in the majority of situations, the simple paired ANOVA is not an appropriate approach.

Given that the test of interaction may be of limited power in the context of an *N*-of-1 trial, because of the relatively small number of period pairs, one should err on the side of caution and in general use the more conservative paired *t*-test. However, in those situations where there is no evidence of extra variation (say when the mean square interaction is less than or equal to the mean square error), then one should consider the comparison based upon the pooled error. We acknowledge that many statisticians may be reluctant to accept our position regarding the appropriateness of use of any parametric methods, including the paired *t*-test. More conservative practitioners, therefore, will restrict themselves to use of nonparametric methods.

Interpreting the Data: Autocorrelation

Large group studies are usually so designed that, within the constraints of the design, the responses of the different patients are statistically independent. On the other hand, data for successive treatment periods for one and the same patient are of the nature of a "time series" and may be serially dependent or "autocorrelated." That is, any individual observation is to some extent a

function of the previous observation. Thus, an individual observation may bear a closer relation to an adjacent observation than to an observation much earlier or later in the series. Serial dependence can occur for instance when there is a natural ebb and flow of the severity of the underlying disease process. The lack of independence of individual outcome assessments over time, if present, is important to recognize since it constitutes a serious threat to the valid application of either parametric or nonparametric methods to *N*-of-1 RCTs.

While one can postulate all kinds of biologic reason why symptom scores could be expected to be correlated from day to day, it is necessary to formally examine many *N*-of-1 trials in similar disease processes to demonstrate empirically the level of dependence. One such investigation has already been reported by Huitema.¹⁶ He reviewed the results of almost 500 *N*-of-1 trials in the behavioral area, and concluded there was very little evidence of important serial dependence. He refers to the widely held belief that autocorrelation is a problem in *N*-of-1 studies as a "myth."

We have conducted a similar analysis in 16 *N*-of-1 trials in patients with a spectrum of medical conditions. The 16 *N*-of-1 RCTs involved repeated outcome assessments during each treatment period, and between three and six pairs of treatment periods. Autocorrelations were estimated from data which had first been "detrended" by subtraction of the period mean to produce residual variation. Serial correlation coefficients can be calculated for pairs of residuals which are one time unit apart (i.e., adjacent in time), two units apart, etc. Even if there is no real correlation between successive raw observations, the subtraction of period means will create slight negative correlation among pairs of residuals within a period. This correlation, equal to $1/(k-1)$ (where k is the number of observations per period), is created by the fact that the sum of residuals within a period must add to zero. One must take this feature into account when examining observed residuals for autocorrelation.

In practice, only the first few autocorrelations need to be investigated. Since one might expect that serial dependence, if present in this type of

data, would be in the form of a first order autoregressive process (highest correlation between adjacent observations), one should be sensitive to the appearance of the characteristic exponentially decaying pattern of autocorrelations. We found almost no evidence of autocorrelation of individual assessments over time and reasonable normality for a variety of composite symptom scores and physiologic outcomes (Table 37.3). Although 22 of 64 of the autocorrelations shown in Table 37.3 are greater than 0.2, only five of these are statistically significant and only two of these are first order correlations. The pattern of autocorrelation is little different from that which would be expected by chance if there were no underlying autocorrelation at all.

While the power of the tests for autocorrelation is limited, the data nevertheless suggest that autocorrelation is the exception rather than the rule. Autocorrelations of 0.2 and higher can alter *p*-values, so that the true *p*-value will be greater than the *p*-value that one observes when one conducts the statistical test.¹⁷ An autocorrelation of 0.2 for instance, will result in a true *p*-value of 0.089 being associated with an observed *p*-value of 0.05. However, even if such autocorrelations were present in the individual data, the autocorrelation of the mean scores from each treatment period (which we advocate using in the paired *t*-test described above) would be substantially lower and unlikely to significantly distort observed *p*-values.

If one wishes to be reasonably secure in the use of parametric or nonparametric methods that assume no serial dependence in an individual

N-of-1 RCT, one can begin by testing for the presence of autocorrelation of the residuals. In our trial of ipratropium, for instance, the first and second order autocorrelations (\pm standard error) for the symptom data were 0.10 (\pm 0.11) and -0.06 (\pm 0.10) respectively. The corresponding figures for the peak flow data were 0.07 (\pm 0.16) and 0.24 (\pm 0.16). These autocorrelations are not statistically significant and their magnitude is small. The results therefore support application of standard parametric or nonparametric analysis to the data. Had significant autocorrelation been found, an analytic method applicable for dealing with autocorrelated data from *N*-of-1 RCTs could have been applied.¹⁸

A summary of our viewpoint regarding autocorrelation is as follows. First, in 16 *N*-of-1 studies we found a consistently small magnitude of autocorrelation. The pattern of the serial dependence observed suggests that most of the autocorrelation was due to the play of chance, rather than a true finding. Although this is a small sample, the results suggest that autocorrelation of large magnitude may be unusual in *N*-of-1 RCTs. Second, Huitema's study of almost 500 single case studies supports the hypothesis that important correlation is unusual in these experiments. Third, the impact of any autocorrelation would be minimized by using mean scores from each period and by the paired design. Fourth, if concern regarding autocorrelation remains, the data from an individual study should be tested to determine the extent of autocorrelation.

Table 37.3. Number of *N*-of-1 RCTs showing varying magnitudes of autocorrelation

Degree of separation	Magnitude of autocorrelation				
	≤ 0.1	$>0.1 - \leq 0.2$	$>0.2 - \leq 0.3$	$>0.3 - \leq 0.4$	>0.4
+ 1	7/16	3/16	2/16	2/16 ^a	2/16 ^a
+ 2	5/16	7/16	2/16	2/16	0/16
+ 3	4/16	8/16	3/16	1/16	0/16
+ 4	5/16	3/16	4/16	3/16 ^b	1/16 ^a

No correlations reach conventional levels of statistical significance unless otherwise stated.

^aOne of these correlations reached conventional levels of statistical significance ($p < 0.05$).

^bTwo of these correlations reached conventional levels of statistical significance ($p < 0.05$).

N-OF-1 RCT: CLINICAL INTERPRETATION OF THE RESULTS

We have touched on the problem of interpreting the results of studies using HRQL measurements in assessing the effectiveness of therapeutic interventions in the “Quality of life” chapter in this book (see Chapter 36). In that chapter we noted that, while conducting classical (multi-subject) RCTs and measuring symptoms on a seven point scale, we have noted that a difference of 0.5 points per question approximates the smallest important difference. This observation was made when the minimal important difference was defined as the smallest difference in score in the domain of interest which patients perceive as a change and which would mandate, in the absence of troublesome side-effects and excessive cost, modification in the patient’s management. While the clinician would participate in the decision regarding modification of management, the definition otherwise focuses on the patient’s experience. This follows from a conceptual or philosophical perspective that sees quality of life as part of an individual’s subjective experience.

Conducting *N*-of-1 RCTs provides an opportunity to define the size of the minimally important difference and other effect sizes. The strategy that we have used in the *N*-of-1 RCTs performed to date could be summarized as follows.¹⁹

All these *N*-of-1 RCTs were designed to examine the efficacy of specific interventions in ameliorating symptoms due to a variety of conditions. The primary outcome measure in each *N*-of-1 RCT was an HRQL questionnaire measuring the severity of symptoms identified by patients as related to their disease and important in their day-to-day life. The operational definition of the minimal important difference that we used was the smallest difference that was important enough that patients would choose to continue indefinitely with the intervention.

We conducted each trial according to the principles outlined in the previous sections. To assess drug efficacy and side-effects, we constructed an individualized questionnaire examining the severity of symptoms identified by patients as part of their disease and as being important in

their day-to-day life. We identified the symptoms to be measured from detailed interviews with patients in which we elicited patients’ experience of the illness or the drug, and what bothered them most. The questionnaires that followed consisted of four to seven items (symptoms), with severity of symptoms measured on a seven point scale (see previously presented response option example). We calculated the difference in the mean score per question between treatment periods for each treatment pair.

To directly assess the patient’s perception of the drug’s effect, we asked the following questions after each pair of treatment periods:

Overall in which of the two periods did you feel better?

1. First period
2. Second period
3. No difference.

If patient expressed a preference on the above question we then asked:

Would you continue drug A indefinitely if it was actually drug A that made you feel better?

1. Yes
2. No

When the patient answered *yes* to the above questions we asked her/him to provide us with the magnitude of the drug effect by asking the following question (global rating):

If it turns out that you felt better during the period in which you were on drug A, we would like you to rate how important the difference between the two periods is to you:

1. Not important
2. Slight importance
3. Some importance, consistent benefit
4. Moderate importance, consistent benefit
5. Much importance, good deal of benefit
6. Very important
7. Great importance.

A global rating score of 0 was assigned if the patient indicated that there was no difference between periods, or if the observed difference was not sufficient to make him or her take the drug indefinitely.

We subsequently examined the relation between patients' subjective assessment of drug efficacy (global rating) and differences in the quality of life questionnaire score in every pair of every *N*-of-1 RCT in which data were available. As different questionnaires included from four to seven questions, the difference in the HRQL score was expressed as the total difference in score divided by the number of questions in a particular questionnaire. The minimal important difference was defined as the difference in the questionnaire score corresponding to a small degree of importance (answers (1)–(3) on the global rating); moderate benefit as differences corresponding to answers (4) or (5); and large benefit as differences corresponding to answers (6) or (7).

To help understand the significance of different changes in questionnaire score, we calculated the mean differences on the HRQL questionnaire score in patients who told us they had experienced small, moderate, and large changes using the global rating. A mean difference of 0.29 points per question in HRQL questionnaire score corresponded to a global rating of a small degree of change—the minimal important difference. Differences of approximately 0.66 points per question corresponded to the moderate difference by the global rating; differences of about 1.09 points per question represented a marked difference.¹⁹ We have found similar results in the setting of quality of life measurement in conventional multipatient studies using a variety of interventions, though the difference corresponding to a minimal important difference has tended to be somewhat higher (around 0.5).^{20–23}

While analyzing the size of minimal important difference in the setting of *N*-of-1 RCTs, we observed large between-patient variability in the changes in symptom questionnaire score, corresponding to varying estimates of drug efficacy. Some portion of this variability is certainly due to the less than perfect validity of the independent standard (in this case, the global rating of drug

efficacy). However, it is likely that patients have different standards about the changes in symptoms that they view as important or trivial. Physiological measures also display the same variability in the clinical significance of a particular change in score or value. On the other hand, establishing the range of changes in score that correspond to small, medium, and large effects across a group of patients undergoing *N*-of-1 RCTs and confirming that these changes conform to the previous estimates²⁴ provided us with information allowing meaningful interpretation of study results and was useful in the planning of future studies.

N-OF-1 RCTs: PRACTICAL ASPECTS

Conducting an *N*-of-1 RCT that incorporates the appropriate safeguards against bias and misinterpretation can be challenging. Usually, the clinician requires collaboration with a pharmacist or pharmacy service. Preparation of placebos identical to the active medication in appearance, taste, and texture is required. Occasionally, pharmaceutical firms can supply such placebos. More often, however, clinicians must collaborate with a local pharmacist to repackage the active medication; if it comes in pill form, it can be crushed and repackaged in capsule form. Identical appearing placebo capsules can be filled with lactose. While somewhat time consuming, preparation of placebo is not technically difficult. Our own average cost for preparing medication for *N*-of-1 studies in which placebos have not been available from the drug's manufacturer has been \$125. The pharmacist is also generally charged with preparing the randomization schedule (which requires nothing more than a "coin toss" for each pair of treatment periods). This allows the clinician, along with the patient, to remain blind to allocation.

The pharmacist may also be helpful in planning the design of the trial through knowledge of the drug's action and the washout period needed, thus helping with decisions about the duration of study periods. The pharmacist can also help monitor compliance and drug absorption. Both pill counts and the drawing of serum drug levels at the end of each treatment period can help establish that the

patient is taking the study medication conscientiously throughout the trial.

Another major issue is the ethics of an *N*-of-1 RCT. Is the conduct of an *N*-of-1 RCT a clinical or a research undertaking? If the former, is it the sort of clinical procedure, analogous to an invasive diagnostic test, that requires written informed consent? We would argue that the *N*-of-1 RCT could, and should be, a part of routine clinical practice. However, like all medical innovations, it may require a period of experimentation and study before being accepted by clinicians, and by ethics committees.

Nevertheless, a number of ethical issues are important. We believe that patients should be fully informed of the nature of the study in which they are participating, and there should be no element of deception in the use of placebos as part of the study. Clinicians should obtain written informed consent (the consent form we use is presented in the Appendix). Patients should be aware that they could terminate the trial at any time without jeopardizing their care, or their relationship with their physician. Finally, followup should be close enough to prevent any important deleterious consequences of institution or withdrawal of therapy.

All these considerations suggest that setting up an *N*-of-1 trial can involve a substantial amount of effort. One must explain the endeavor fully to the patient, obtain informed consent, construct the outcome questionnaire, negotiate with the pharmacist regarding preparation of medication, carry out the study, and conduct the analysis. These challenges may seem formidable to the busy clinician. Nevertheless, with effort, they can be managed.^{25–27}

N-OF-1 RCTs: EXPLORING NEW DRUG USE

The Conventional Approach

In the preceding sections, we described the application of *N*-of-1 RCTs in deciding about optimal therapy in individual patients. The concept of *N*-of-1 RCTs, however, could also be used to draw more general conclusions regarding drug efficacy. The main role for this methodology is

likely in the early, premarketing stages of drug development,²⁸ but the same principles may apply in the postmarketing phase while exploring new areas of possible drug utilization—using different doses, examining the effects in different populations, or investigating previously unrecognized beneficial actions or side effects.

Let us use as an example the hypothetical observation that a long used antidepressant medication (drug A) improves itching associated with renal failure. Although this example is hypothetical, it is not inconceivable; recall amantadine's use in Parkinson's disease, β -blockers in familial tremor, or amitriptyline in fibrositis, chronic pain, or diabetic neuropathy. Now, the investigators wish to compare drug A with currently used modes of therapy. In large sample parallel group trials, patients are assigned at random to drug A or an alternative drug (say, an antihistamine). The treatment groups are followed and the treatment target monitored. These trials are the standard approach to establishing drug efficacy, and to persuading regulatory agencies that a new medication should be placed on the market or a new indication for an old medication accepted.

There are three major hurdles that need to be overcome before such large sample parallel group studies of the efficacy and safety of a drug can be undertaken. First, investigators must determine whether the drug shows sufficient promise to justify the initiation of a large clinical research program. Second, they must define the patient population to be studied. Third, they must establish the dose regimen to be used in the major trials.

These decisions are generally based on findings from early clinical safety, tolerance, pharmacology, and drug disposition studies in healthy volunteers and patients, augmented by ideas gained from initial small scale efficacy studies. These efficacy studies are often open, and use baseline status or a historical reference group as a control. Such studies tend to yield anecdotal information of questionable validity. Investigators may use classical double blind randomized parallel groups studies in the early exploration of drug properties, but their small sample size results in

findings that leave considerable uncertainty about the answers to the three key questions described above.

Thus, when designing the first large sample efficacy study, investigators face difficult decisions concerning both dose regimen and sample selection. They may gamble on a single dose (with a possibility of suboptimal efficacy with too small a dose or excessive side-effects with too large a dose), or take a safer approach that includes two or more different regimens for comparison. At the same time, they may hazard a guess at a suitable homogeneous target population of patients, or take a more conservative approach that includes a heterogeneous (possibly stratified) population.

If the investigators decide to gamble or guess and turn out to be wrong, the large sample efficacy study misfires. Nevertheless, the choice for gambling and guessing is frequently made. The reason is that, even with a well defined, homogeneous patient population, a parallel groups study of a single dose of a new drug often requires large numbers of patients for adequate statistical power. The extra numbers required for a heterogeneous patient population and/or several different dosage regimens might well be considered prohibitive.

Even if, through good luck or sound judgment, the first large sample efficacy study is successful and the study population is clearly shown to benefit from the selected dosage of the new drug, important questions are likely to remain. Would an equivalent benefit have been obtained at lower doses, or would there be additional benefit with higher doses? Will the use of lower dose allow preservation of drug benefit with reduced side effects? Are there subgroups of patients who are particularly responsive or resistant to the new drug?

Several successive rounds of large sample parallel group studies may well fail to provide clearcut answers. One reason is that in such studies, there is usually much uncontrolled variation among patients. The result of this variation is that the determination of the profiles of responsive and drug resistant subpopulations of patients is hindered and at times rendered impossible.

The Role of *N*-of-1 RCTs

N-of-1 trials share many features in common with the traditional crossover trials used in drug evaluation. The fundamental difference between the *N*-of-1 approach and traditional crossover trials is their primary purpose: *N*-of-1 trials attempt to establish effects in an individual, whereas crossover trials attempt to establish effects in a group. As a secondary goal, one may use a crossover trial to examine individual responses. Similarly, one may analyze a series of *N*-of-1 trials with a similar design as a multiple crossover trial. However, the *N*-of-1 trial will be designed so that individual effects can be reliably detected; the crossover trial will be designed so that individual estimates of response are imprecisely estimated but the magnitude of the average group effect can be efficiently determined.

In some therapeutic indications (including our hypothetical example of itch), the problems faced by investigators involved in drug development may be overcome by including in the program a short series of carefully designed *N*-of-1 RCTs. These studies will permit the reliable identification of responders and nonresponders and an estimate of the proportion of patients in each category. They may also make it possible to determine the optimal dosage regimen for individual patients. The availability of this type of information makes the design of large scale parallel group studies less problematic.

In relating the foregoing example to the general use of *N*-of-1 RCTs in drug development, we will deal in turn with the following six issues: the role of an open run-in period, determining the rapidity of onset of drug action, optimizing dose, measuring outcome, assessing potential drug impact, and predicting response.

The Role of an Open Run-in Period

To increase the efficiency of generating data in drug development (or in the exploration of a new indication), one could argue for first conducting formal *N*-of-1 RCTs only among patients who showed an apparent benefit during an open run-in period. Then, only responders in *N*-of-1 trials

would be included in subsequent large sample parallel group studies. In open therapeutic trials, potential biases, including patient and physician expectations and the placebo effect, generally favor a conclusion that the new treatment is beneficial. Thus, a false positive conclusion (concluding the drug works better when it does not) is far more likely than a false negative conclusion (concluding the drug does not work when in fact it does). Accordingly, it would be reasonable to conduct formal *N*-of-1 studies only on the subgroup of patients with apparent benefit from a new drug in open trials. Obviously, any estimates of the degree to which patients prefer drug A over drug B then have to take into account that only potential respondents were included in the formal evaluation.

Determining the Rapidity of Onset and Termination of Action

Conducting visual and statistical analyses of the data generated while obtaining numerous measurements during each treatment period in an *N*-of-1 RCT may help in determining the rapidity of the onset and termination of the treatment effect. One may incorporate this knowledge into the design of subsequent *N*-of-1 RCTs and large sample parallel group studies may then incorporate this knowledge. Precise information about the rapidity of action may also add to the understanding of the drug's mechanism of action—a drug exerting its action within minutes of administration and another drug starting to act only after days of use are likely to have different biological modes of action.

Optimizing Dose

When one does not know the optimal dose of a drug, one option is to allow open, unblinded dose titration to obtain a response in the individual patient. One of many alternative approaches would be to conduct an *N*-of-1 RCT, beginning with the lowest plausible effective dose of the new drug. If the first pair or set of pairs showed superiority of drug B (the accepted method of treatment), a higher dose of drug A could be used

in the next pair. The process could continue until side-effects appeared, the highest acceptable dose of drug A was reached, or a difference in favor of drug A emerged. This last observation could be confirmed by conducting additional pairs of treatment on the apparently favorable dose of drug A. This approach would not only help determine the optimal dose, but would reveal whether this optimal dose differed among different patients, an issue that would be very difficult to elucidate using parallel group studies. In addition, it would be possible to modify the doses used after only a few *N*-of-1 trials, if a high incidence of toxicity (use lower doses) or a low incidence of response (use higher doses) were found.

Measurement of Outcome

In the initial study of a drug in a new setting, investigators may be uncertain about the outcomes upon which to focus. This is particularly true if the primary outcomes relate to patients' symptoms, and if the condition being treated (or the treatment) results in a spectrum of problems. For example, an antidepressant used for itchiness may have differential impact on the primary complaint, mood, and a number of possible side-effects, including problems with vision, dry mouth, and constipation (some of which may be idiosyncratic to people with renal failure). These differences may become apparent in the initial positive *N*-of-1 RCTs, giving the investigator an opportunity to shift the focus of outcome measurement to the areas most likely to show benefit or to important side-effects.

Assessing Potential Drug Impact

Once a number of *N*-of-1 trials have been conducted, one is in a position to evaluate the potential impact of the medication. If only one out of 20 patients has an *N*-of-1 RCT that shows drug A relatively superior, the drug may not be worth further development; if 15 out of 20 patients favor a new drug, one clearly has an important new treatment. Between such extremes, the decision concerning further study of the drug will depend on factors such as the prevalence of the condition

being treated, its associated morbidity, the expense and toxicity of the treatment, and the availability of other effective treatments. If a condition is of major importance to a relatively large population of patients, and if the drug (at the doses used) is inexpensive and nontoxic (both of which would be true for the use of amitriptyline in renal failure induced itchiness), a 25% response rate likely suggests an important role for the medication. In a condition that results in severe morbidity and for which there is no other treatment, the use of an inexpensive and nontoxic drug might be worth exploring and using even if only a small proportion of patients gained a clinically important benefit.

If initial *N*-of-1 RCTs suggest further study is warranted, the results can help in planning subsequent investigations. For example, sample size for a parallel group study can be informed by prior *N*-of-1 RCTs which provide accurate information concerning both within-person variability over time and heterogeneity of treatment response. The lower the response rate in preceding *N*-of-1 RCTs, the larger the sample size required in subsequent parallel group designs to detect the desired clinically important difference.

Predicting Response

N-of-1 RCTs can also help determine eligibility criteria for subsequent studies. The precise identification of responders and nonresponders allows powerful examination of predictors of response. If there is very little overlap between responders and nonresponders (for example, if virtually all people on peritoneal dialysis respond and all those on hemodialysis do not), a small number of *N*-of-1 RCTs will allow identification of variables associated with response. If a larger number of *N*-of-1 RCTs have been completed, weaker predictors may also be identified. If the number of trials is large enough, one can use logistic regression methods to determine the independent contribution of a set of variables in differentiating responders from nonresponders.

Identifying variables associated with response is important for clinicians in deciding when to use a drug. In addition, the ability of *N*-of-1 RCTs to define responders precisely may provide a solution

to one of the major dilemmas facing those exploring the use of a drug in new situation: choosing the population for the first large sample parallel group RCT.

In summary, *N*-of-1 RCTs have an important potential role to play in the development of new drugs or in explorations of the use of an old drug in a new clinical situation. Information regarding rapidity of onset and termination of drug action, the optimal dose, the outcomes on which to focus, and predictors of response may be obtained most efficiently using *N*-of-1 RCTs. The ultimate impact of a medication can be thus assessed early in the process of clinical testing.

THE FUTURE

We have previously reported our experience with the use of *N*-of-1 methodology and found that the method is able to aid in resolution of difficult clinical dilemmas; that based on the results of *N*-of-1 trials treatment frequently changes; and that, with long-term followup, physicians continue to follow the conclusions based on the *N*-of-1 RCT results.⁷ In addition to Larson's work which we have already cited, other clinical groups have reported on their experience with *N*-of-1 RCTs, generally confirming the feasibility and usefulness of the approach.^{29,30}

These reports do not definitively answer the question about whether patients who undergo *N*-of-1 RCTs are better off than those whose treatment regimen is determined by conventional methods. The most rigorous test of the usefulness of *N*-of-1 RCTs would be a randomized trial. Two such trials, in which investigators randomized patients to conventional care or to undergo *N*-of-1 RCTs, have addressed the impact of *N*-of-1 RCTs.^{31,32} Both were conducted by the same group of investigators, and both studied the use of theophylline in patients with chronic airflow limitation. The investigators found that while using *N*-of-1 RCTs did not affect patients' quality of life or functional status, of patients initially on theophylline, fewer in the *N*-of-1 RCT groups ended up receiving the drug in the long term. Thus, *N*-of-1 RCTs saved patients the expense,

inconvenience, and potential toxicity of useless long term theophylline therapy.

While confirming the potential of *N*-of-1 RCTs, groups with extensive experience with *N*-of-1 RCTs have noted the time and effort required. It is unlikely that full implementation of *N*-of-1 RCTs will become a major part of clinical practice. Clinicians can, however, incorporate many of the key principles of *N*-of-1 RCTs into their practice without adopting the full rigor of the approach presented here. Medication can be repeatedly withdrawn and reintroduced in an open or

unmasked fashion. Symptoms and physical findings can be carefully quantified. However, without the additional feature of double blinding, both the placebo effect and physician and patient expectations can still bias the results.

In summary, the *N*-of-1 approach clearly has potential for improving the quality of medical care and the judicious use of expensive and potentially toxic medication in patients with chronic disease. Using the guidelines offered here, we believe that clinicians will find the conduct of *N*-of-1 RCTs feasible, highly informative, and stimulating.

APPENDIX: CONSENT FORM FOR *N*-OF-1 RANDOMIZED TRIALS

Often, doctors and patients try to decide what treatment is best by starting a particular drug and seeing if it makes the patient feel better. Unfortunately, the results of such 'therapeutic trials', as we call them, are often misleading. The patient may have gotten better anyway, even without any medication. Or, the doctor and the patient may be so optimistic that they may misinterpret the results of the therapeutic trial. Finally, people often feel better when they are taking new medication even when it doesn't have any specific activity against their illness (the placebo effect), and this may also lead to a misleading interpretation of the value of the new treatment.

We believe we have developed a better way of conducting therapeutic trials. In this new approach, the patient alternates between taking active drug and placebo. The placebo is a medication that looks identical, but doesn't contain the active ingredient. While taking the active and placebo medication the patient keeps careful track of how he or she is feeling. Neither the doctor nor the patient knows when the patient is taking the active drug and when the patient is taking the placebo. After several periods on both active and placebo medication the 'code' is broken and the patient and the doctor then look at the results. Together, they decide whether the drug is of any benefit and should be continued.

We think that it would help you to take part in one of these therapeutic trials of [NAME OF DRUG]. We will conduct a number of pairs of periods. Each period will be [DURATION OF PERIOD]. During one period of each pair you will be taking the active treatment, and during the other you will be using the placebo. If at any time during the study you are feeling worse we can consider that treatment period at an end and can go on to the next treatment. Therefore, if you begin to feel worse, just call my office at [INSERT NUMBER] and I will get in touch with you.

If you don't think this new way of conducting a therapeutic trial is a good idea for you, we will try the new drug in the usual way. Your decision will not interfere with your treatment in any way. You can decide to stop the trial any time along the way, and this will not interfere with your treatment either. All information we collect during the trial will remain confidential.

PATIENT SIGNATURE _____

WITNESS SIGNATURE _____

PHYSICIAN SIGNATURE _____

DATE _____

REFERENCES

1. Guyatt GH, Sackett DL, Cook DJ for the Evidence-based Medicine Working Group. Users' guides to the medical literature: II. How to use an article about therapy or prevention: Part A. Are the results of the study valid? *J Am Med Assoc* 1993; **270**: 2598-601.
2. Dans AL, Dans LF, Guyatt GH, Richardson S, for the Evidence-based Medicine Working Group. User's guides to the medical literature: XIV. How

- to decide on the applicability of clinical trial results to your patient. *J Am Med Assoc* 1998; **279**: 545–9.
3. Guyatt GH, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy: randomized trials in individual patients. *New Engl J Med* 1986; **314**: 889–92.
 4. Kratchwill TR. *Single Subject Research: Strategies for Evaluating Change*. Orlando, FL: Academic, 1978.
 5. Kazdin AE. *Single-Case Research Designs: Methods for Clinical and Applied Settings*. New York: Oxford University Press, 1982.
 6. Guyatt GH, Sackett DL, Adachi JD, et al. A clinician's guide for conducting randomized trials in individual patients. *Can Med Assoc J* 1988; **139**: 497–503.
 7. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom R, Adachi JD, Newhouse MT. Clinical usefulness of the *N* of 1 randomized control trials: three year experience. *Ann Intern Med* 1990; **112**: 293–9.
 8. Jaeschke R, Adachi JD, Guyatt GH, Keller JL, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 randomized controlled trials. *J Rheumatol* 1991; **18**: 226S–33S.
 9. Patel A, Jaeschke R, Guyatt G, Newhouse MT, Keller J. Clinical usefulness of *N*-of-1 RCTs in patients with chronic airflow limitation. *Am Rev Respir Dis* 1991; **144**: 962–4.
 10. Radack K, Deck C. Beta-adrenergic blocker therapy does not worsen intermittent claudication in subjects with peripheral arterial disease. A meta-analysis of randomized controlled trials. *Arch Intern Med* 1991; **151**: 1769–76.
 11. Jaeschke R, Singer J, Guyatt GH. A comparison of 7 point and visual analogue scales: data from a randomized trial. *Control Clin Trials* 1990; **11**: 43–51.
 12. DeProspero A, Cohen S. Inconsistent visual analysis of intrasubject data. *J Appl Behav Anal* 1979; **12**: 573–9.
 13. Beyer WH, ed., *CRC Handbook Of Tables For Probability And Statistics*. 2nd edn. Cleveland, OH: Chemical Research Company, 1968.
 14. Conover WJ. *Practical Nonparametric Statistics*. New York: Wiley, 1971; 121–6, 206–15.
 15. Edgington ES. *Randomization Tests*. 2nd edn. New York: Dekker, 1987.
 16. Huitema BE. Autocorrelation in applied behavior analysis: a myth. *Behav Assessment* 1985; **7**: 107–18.
 17. Gastwirth JL, Bubin H. Effect of dependence on level of significance of a one sample test. *J Am Stat Assoc* 1971; **66**: 816–20.
 18. Rochon J. A statistical model for the “*N*-of-1” study. *J Clin Epidemiol* 1990; **43**: 499–508.
 19. Jaeschke R, Guyatt GH, Keller J, Singer J. Interpreting changes in quality-of-life score in *N* of 1 randomized trials. *Control Clin Trials* 1991; **12** (suppl): 226S–33S.
 20. 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.
 21. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994; **47**: 81–7.
 22. Juniper EF, Guyatt GH, Griffith LE, Ferrie PJ. Interpretation of rhinoconjunctivitis quality of life questionnaire data. *J Allergy Clin Immunol*. 1996; **98**: 843–5.
 23. Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in children with asthma. *Qual Life Res* 1996; **5**: 35–46.
 24. Jaeschke R, Singer J, Guyatt G. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; **10**: 407–15.
 25. Keller JL, Guyatt GH, Roberts RS, Adachi JD, Rosenbloom D. An *N*-of-1 service: applying the scientific method in clinical practice. *Scand J Gastroenterol* 1988; **23**: 22–9.
 26. Larson EB. *N*-of-1 clinical trials. A technique for improving medical therapeutics. *West J Med* 1990; **152**: 52–6.
 27. Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a two-year period. *J Am Med Assoc* 1993; **270**: 2708–12.
 28. Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS. *N*-of-1 trials for investigating new drugs. *Control Clin Trials* 1990; **11**: 88–100.
 29. Menard J, Serrurier D, Bautier P, et al. Crossover design to test antihypertensive drugs with self-recorded blood pressure. *Hypertension* 1988; **11**: 153–9.
 30. Johannessen T. Controlled trials in single subjects. 1. Value in clinical medicine. *Br Med J* 1991; **303**: 173–4.
 31. Mahon J, Laupacis A, Donner A, Wood T. Randomised study of *n* of 1 trials versus standard practice. *Br Med J* 1996; **312**: 1069–74.
 32. Mahon J, Laupacis A, Hodder et al. Theophylline for irreversible chronic airflow limitation: a randomized study comparing *n* of 1 trials to standard practice. *Chest* 1999; **115**: 38–48.

38

The Use of Meta-analysis in Pharmacoepidemiology

JESSE A. BERLIN

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

INTRODUCTION

Meta-analysis has been defined as “the statistical analysis of a collection of analytic results for the purpose of integrating the findings.”¹ Other definitions have included qualitative, as well as quantitative, analyses.² Meta-analysis is used to identify sources of variation among study findings and, when appropriate, to provide an overall measure of effect as a summary of those findings.³ While epidemiologists have been cautious about adopting meta-analysis,^{4–6} the need to make the most efficient and intelligent use of existing data prior to (or instead of) embarking on a large, primary data collection effort, has dictated a progressively more accepting approach.^{6–12}

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.¹³ The distinguishing feature of meta-analysis, as opposed to the usual qualitative literature review, is its systematic, structured, and presumably objective presentation and analysis of available data. The traditional

review has been increasingly recognized as being subjective.^{13–16} With the support of leading scientists¹⁷ and journal editors,¹⁸ there has been growing acceptance of the concept that the literature review can be approached as a more rigorous scientific endeavor, specifically, an observational study with the same requirements for planning, prespecification of definitions, use of eligibility definitions, etc., as any other observational study. In recent years, the terms *research synthesis* and *systematic review* have been used to describe the structured review process, in general, while *meta-analysis* has been reserved for the quantitative aspects of the process. For the purposes of this chapter, we shall use *meta-analysis* in the more general sense. Meta-analysis provides the conceptual and quantitative framework for such rigorous literature reviews; similar measures from comparable studies are tabulated systematically and the effect measures are combined when appropriate.

Several activities may be included under the above definition of meta-analysis. Perhaps the most popular conception of meta-analysis, for

most clinically oriented researchers, is the summary of a group of randomized clinical trials dealing with a particular therapy for a particular disease. An example of this approach would be a study that examined the effects of aspirin following myocardial infarction. Typically, this type of meta-analysis would present an overall measure of the efficacy of treatment, 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, when results have been conflicting, attempt to uncover the sources of the disagreements.¹⁹

More recently, meta-analyses of nonexperimental epidemiologic studies have been performed^{20–28} and articles have been written describing the methodologic considerations specific to those meta-analyses.^{3,6,11–13,29–38} In general, both 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. Given the greater diversity of designs of nonexperimental studies, it is logical to find more disagreement among nonexperimental studies than among randomized trials.

This chapter summarizes many of the major conceptual and methodologic issues surrounding meta-analysis and offers the views of one meta-analyst about possible avenues for future research in this field.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

There are a number of reasons why a pharmacoepidemiologist might be interested in conducting a meta-analysis. These include the study of uncommon negative outcomes of therapies (adverse drug reactions) 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.

The investigation of adverse effects has been a recurring theme throughout this book, as it is a major focus of pharmacoepidemiology. It is most often, but not always, pursued through nonexperimental studies. The difficulties in studying these events have also been detailed throughout this book. One major challenge involves obtaining information on adverse reactions that is unconfounded by indication (see Chapter 43). These adverse events often occur only rarely, making their evaluation still more difficult. The results of nonexperimental studies of whether such events are associated with a particular drug may be conflicting, leaving a confusing picture for the practicing clinician and the policy makers to interpret. Meta-analysis, by combining results from many randomized studies, can address the problem of rare events and rectify the associated lack of adequate statistical power in a setting free of the confounding and bias of nonexperimental studies. For example, Chalmers and colleagues used meta-analysis of randomized clinical trials to explore possible gastrointestinal side-effects of nonsteroidal anti-inflammatory drugs (NSAIDs).³⁹ These studies individually had almost no power to detect any association between NSAIDs and adverse gastrointestinal outcomes, but collectively the number of subjects was adequate both to show some important associations and to show the rarity of most complications. The details of this NSAID meta-analysis will be presented later in this chapter.

When reports of several investigations of a specific adverse drug reaction disagree, whether randomized or nonexperimental in design, meta-analysis can also be used to help resolve these disagreements. These disagreements among studies may arise from differences in the choice of endpoints, the exact definition of exposure, the eligibility criteria for study subjects, the methods

of obtaining information, other differences in protocols, or a host of other reasons possibly related to the quality of the studies. While it is not possible to produce a definitive answer to every research question, the exploration of the reasons for heterogeneity among studies' results may at least provide valuable guidance concerning the design of future studies.

The exploration of subgroups of patients in whom therapy may be more or less effective is a controversial question in individual randomized trials. Most trials are not designed with sample sizes adequate to address efficacy in subgroups. The finding of statistically significant differences between the effects of therapy in different subgroups, particularly when those groups were not defined prospectively, raises the question of whether those are spurious findings. Conversely, the lack of statistical significance for clinically important differences between prospectively defined subgroups can often be attributed to a lack of statistical power. Such clinically meaningful but statistically nonsignificant findings are difficult to interpret. Meta-analysis can be used to explore these questions with improved statistical power.

The use of meta-analysis in the approval process for new drugs represents another potential application, although experience in this area is as yet rather limited. However, many of the methodologic issues arising in the context of new drug approval also arise in the investigation of new indications for pharmaceutical products that have previously been approved for other purposes. For some therapies, such as streptokinase in the treatment of myocardial infarction, meta-analysis could have been used to summarize evidence prior to embarking on a very large scale, multicenter, randomized trial.⁴⁰

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY 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 seems particularly prone to the "garbage in, garbage out" phenomenon. Combining a group of poorly done studies can produce a precise summary result built on a very weak foundation. This apparent precision may lend undue credibility to a result that truly should not be used as a basis for formulating clinical or policy strategies.⁵ However, if the quality judgment is subtly influenced by the direction or magnitude of the findings of the study, excluding studies based on such a subjective judgment about their quality could open the meta-analytic process to a serious form of bias.

COMBINABILITY OF STUDIES

Clearly, no one would suggest combining studies that are so diverse that a summary would be nonsensical. For example, one would not combine studies of aspirin in the primary prevention of coronary heart disease with studies of aspirin given after myocardial infarction. Beyond obvious examples like this, however, the choices may not be so clear. Should studies with different patient populations be combined? How different can those populations be before it becomes unacceptable to combine the studies? Should nonrandomized studies be combined with randomized studies? Should nonrandomized studies ever be used in a meta-analysis? These are questions that cannot be answered without generating some controversy.

PUBLICATION BIAS

Unpublished material cannot be retrieved by literature searches and is likely to be difficult to find referenced in published articles. *Publication bias* occurs when study results are not published, or their publication is delayed, because of the results.⁴¹⁻⁵⁰ The usual pattern is that statistically significant results are published more easily than

nonsignificant results, although this bias may not be as severe for randomized studies as it is for nonrandomized studies.^{42,51} While one could simply decide not to include unpublished studies in a meta-analysis, since those data have often not been peer reviewed,⁵² unpublished data can represent a large proportion of all available data.⁵³ If the results of unpublished studies are systematically different from those of published studies, particularly with respect to the magnitude and/or direction of the findings, their omission from a meta-analysis would yield a biased summary estimate (assuming that the quality of the unpublished studies is at least equal to the quality of the published studies).

Publication bias is a potentially serious limitation to any meta-analysis. The retrospective identification of completed unpublished trials is clearly possible⁵³ in some instances, but generally is not practical. One study⁵⁴ used a survey of investigators to attempt to identify unpublished studies. The authors surveyed 42 000 obstetricians and pediatricians, asking whether they had participated in any unpublished trials completed more than two years previously, i.e., during the period prior to the end of 1984. They identified only 18 such studies, despite an overall response rate of 94% to their survey.

Other forms of bias, related to publication bias, have also been identified.⁴⁷ These include reference bias, i.e., preferential citation of significant findings,⁵⁵ language bias, i.e., exclusion of studies in languages other than English,^{56,57} and bias related to source of funding.^{42,43,58-60}

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.⁶¹ In the meta-analysis of gastrointestinal side-effects of NSAIDs mentioned above³⁹ and described more fully below, Chalmers and colleagues examined over 500 randomized studies. They measured the agreement of different individuals when reading Methods sections of papers that had been masked as to their source and the results. There were disagreements on

10–20% of items, which had to be resolved in conference with a third person. These disagreements arose from errors on the part of the reader and from lack of clarity of the presentation of material in the original articles. Whatever its source, when such variability exists, the opportunity for observer bias may exist as well.⁶¹

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 instances⁵² 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. For example, differences between meta-analyses of the same topic in the acceptance and rejection of papers did not always lead to differences in conclusions.

More generally, the acceptance or rejection of different sets of studies can drastically change conclusions. This is illustrated by several meta-analyses of whether or not corticosteroid drugs cause peptic ulcer. The first published paper argued that corticosteroids did not cause peptic ulcer, because the *p*-value for the meta-analysis was only 0.07.⁶² Five years later, a second analysis, by a second set of authors, included a larger number of studies and found evidence for an association with a *p*-value of less than 0.001.⁶³ Re-analysis of the data from the second meta-analysis by the authors of the first meta-analysis, with the addition of several more studies, gave a *p*-value of 0.40.⁶⁴ Another meta-analysis done by the second set of authors gave a revised *p*-value of 0.01.⁶⁵ Despite efforts to make meta-analysis an objective, reproducible activity, there is evidently some judgment involved.

In a separate commentary,⁶⁶ DerSimonian reanalyzed data from one meta-analysis and one clinical review of parenteral nutrition with branched-chain amino acids in hepatic encephalopathy. She pointed to differences in the data

extracted by the two sets of authors^{67,68} for the same endpoints from the same original papers. When combined statistically, the data extracted by the two sets of authors led to substantively different conclusions about the efficacy of therapy.

CURRENTLY AVAILABLE SOLUTIONS

This section will first present the general principles of meta-analysis and a general framework for the methods typically employed in a meta-analysis. Much of this material has been presented in review articles in major clinical journals,^{7,9,10} so only the most important points will be highlighted here. In the second part of this section, specific solutions to the methodologic issues raised in the previous section are presented. Finally, case studies of applications that should be of interest to pharmacoepidemiologists will be presented, illustrating approaches to some of the clinical and methodologic problems raised earlier.

STEPS INVOLVED IN PERFORMING A META-ANALYSIS (SEE TABLE 38.1)

Define the Purpose

While this is an obvious component of any research, it is particularly important to define precisely the primary and secondary objectives of a meta-analysis. The important primary question might be "Are NSAIDs associated with an increased risk of gastrointestinal side effects?". Another might be "Are corticosteroids effective in the treatment of alcoholic hepatitis?". Secondary objectives might include the identification of subgroups in which a treatment appears to be

uniquely more or less effective. For NSAIDs, estimating the absolute risk difference (and, thus, the public health implications) as well as the relative risk (and, thus, the etiologic implications) might be a secondary objective.

Perform the Literature Search

While computerized searches of the literature can facilitate the retrieval of all relevant published studies, these searches are not always reliable. Several studies have examined problems with the use of electronic searches.⁶⁹⁻⁷¹ Use of search terms that are too nonspecific can result in large numbers of mostly irrelevant citations that need to be reviewed to determine relevance. Use of too many restrictions can result in missing a substantial number of relevant publications. For example, in preparing for meta-analyses of neonatal hyperbilirubinemia, MEDLINE was searched for relevant clinical trials.⁶⁹ A search by a trained librarian identified only 29% of known trials in the Oxford Database of Perinatal Trials.⁷² It is generally suggested that a professional librarian with training and experience in searches of clinical topics be consulted, although, as just cited, even a trained librarian may not perform perfectly. Other methods of searching, such as review of the reference sections of retrieved publications found to be relevant, and hand searches of relevant journals, are also recommended.

Establish Inclusion/Exclusion Criteria

A set of rules for including and excluding studies from the meta-analysis should be defined during the planning stage of the meta-analysis and should be based on the specific hypotheses being tested in the analysis. One might, for example, wish to limit consideration to randomized studies with more than some minimum number of patients. In a meta-analysis of epidemiologic studies, one might wish to include studies of incident cases only, excluding studies of prevalent cases, assuming that the relationship between exposure and outcome could be different in the two types of study. Practical considerations may, of course, force changes in the inclusion criteria. For example,

Table 38.1. 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 recommendation

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. If broad inclusion criteria are established, then a broad, and perhaps more generalizable, hypothesis may be tested. The use of broad entry criteria also permits the examination of the effects of research design on outcome (e.g., do randomized and nonrandomized studies tend to show different effects of therapy?) or the exploration of subgroup effects. In the example of aspirin administered following myocardial infarction, restriction of the meta-analysis to studies using more than a certain dose of aspirin would not permit an exploratory, cross-study comparison of dose-response effects, which might prove illuminating.

A key point is that exclusion criteria should be based on *a priori* considerations of design of the original studies and completeness of the reports and specifically 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.¹¹ While that may seem obvious, the temptation to try to justify such exclusions on a *post hoc* basis may be strong, particularly when a clinically plausible basis for the exclusion can be found. Such exclusions made after having seen the data, and the effect of individual studies on the pooled result, may form the basis for legitimate sensitivity analyses (comparing pooled results with and without that particular study included), but should not be viewed as primary exclusion criteria.

Another important note is that studies may often generate more than one published paper. For example, later reports might update analyses previously published, or might report on outcomes not addressed in earlier papers. It is essential, for two reasons, that only one report on the same patients be accepted into the meta-analysis. First, the validity of the statistical methods depends on the assumption that the different studies represent different groups of individuals. Second, the inclusion of a study more than once would assign undue

weight to that study in the summary measure. A caution is that it is not always obvious that the same patients have been described in two different publications. Contacting the authors may be of some help in determining whether there is duplication, although some authors may perceive the inquiry as questioning their academic integrity. The issue of multiple publications based on the same study has been addressed in more detail by Huston and Moher.⁷³

Collect the Data

When the relevant studies have been identified and retrieved, the important information regarding study design and outcome needs to be extracted. Typically, data abstraction forms are developed, pilot tested on a few articles, and revised as needed. As in any research, it is necessary to strike a balance between the completeness of the information abstracted and the amount of time needed to extract that information. Careful specification in the protocol for the meta-analysis of the design features and patient characteristics that will be of clinical or academic interest may help avoid over- or undercollecting information. It is generally advisable, when possible, to collect raw data on outcome measures, e.g., numbers treated and number of events in each group, rather than derived measures such as odds ratios, which may not be the outcome measures of interest in the meta-analysis or may have been calculated incorrectly by the original authors.

Many articles on "how to do a meta-analysis" (e.g.,^{9,10}) recommend that the meta-analyst assess the quality of the studies being considered in a meta-analysis. One might wish to use a measure of study quality as part of the weight assigned to each study in the analysis, as an exclusion criterion (e.g., excluding studies with quality scores below some arbitrary threshold), or as a stratification factor allowing the separate estimation of effects for good quality and poor quality studies.^{74,75} Chalmers and colleagues have developed a quality assessment scoring system for randomized trials.⁷⁶ Several groups have opted for other, far shorter and simpler, systems (e.g.,⁷⁷⁻⁸⁰) Issues related to quality scoring have been discussed more generally

by Moher and colleagues,⁸¹ and an annotated checklist of quality scoring systems is available.⁸² Most of these systems were proposed as very general systems that could be applied to clinical trials covering a wide range of therapies and endpoints. Scoring systems designed for epidemiologic studies have been developed as well, in the context of evaluating studies of specific exposure–disease relationships (e.g.,²¹).

The argument has been made, however, that general scoring systems are arbitrary in their assignment of weights to particular aspects of study design, and that such systems risk losing information, and can even be misleading.^{38,83,84} Juni and colleagues,⁸⁴ for example, examined studies comparing low molecular weight heparin with standard heparin with respect to prevention of postoperative thrombosis. They used 25 different quality assessment scales to identify high quality trials. For six scales, the studies identified as being of high quality showed little to no benefit of low molecular weight heparin, while for seven scales, the “high quality” studies showed a significant advantage of low molecular weight heparin. This apparent contradiction raises questions about the validity of such scales as methods for stratifying studies.

Thus, in a given meta-analysis, one might wish to examine *specific* aspects of study design that are unique to that clinical or statistical situation.^{38,83–85} For example, Schulz and colleagues⁸⁵ found that trials in which the concealment of randomized allocation was inadequate on average produced larger estimates of treatment effects compared with trials in which allocation was adequately concealed. This specific finding was not detected when these same authors looked for an overall association between quality score and treatment effect. In the analysis of low molecular weight heparin, Juni and colleagues⁸⁴ found that studies with unmasked outcome assessment showed larger, and presumably biased, benefits of low molecular weight heparin than studies using masked assessment of outcome.

Two procedural recommendations have been made regarding the actual techniques for data extraction. One is that studies should be read independently by two readers. The justification for

this comes from meta-analyses in which modest but important inter-reader variability has been demonstrated.^{39,61} A second recommendation is that readers be masked to certain information in studies, such as the identity of the authors and the institutions at which a study was conducted, and masked to the specific treatment assignments.⁵² While masking has a high degree of intuitive appeal, the effectiveness of masking in avoiding bias has not been demonstrated. Only one randomized trial examines the issue of the effect of masking on the results of meta-analyses.⁸⁶ This study compared the results of the same meta-analyses performed independently by separate teams of meta-analysts, with one team masked and the other unmasked. The masked and unmasked teams produced nearly identical results on a series of five meta-analyses, lending little support to the need for masking.

Perform Statistical Analyses

In most situations, the statistical methods for the actual combination of results across studies are fairly straightforward. 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.⁸⁷ 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.^{29,31,87} Other approaches include inverse variance weighted averages of study-specific estimates of multivariate adjusted relative risks^{29,87} and exact stratified odds ratios.⁸⁸ One popular method, sometimes called the “one-step” method, which is similar to the Mantel–Haenszel method,⁸⁹ has been shown to be biased under some circumstances. Since this method offers no clear advantage over the Mantel–Haenszel method, which is more robust, the Mantel–Haenszel method may be preferable in many circumstances.³¹ Choices are never simple, however. In a simulation study, Deeks and colleagues⁹⁰ showed, at an international meeting, 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.

One basic principle in most analytic approaches is that the comparisons between treated (exposed) and untreated (unexposed) patients are typically made within a study prior to combination across studies. In the combination of randomized trial results, this amounts to preserving the randomization within each study prior to combination. In all of the procedures developed for stratified data, “study” plays the role of the stratifying variable. In general, more weight is assigned to large studies than to small studies because of the increased precision of larger studies. A second basic principle to note is that these methods generally assume that the studies are all estimating a single, common effect, e.g., a common odds ratio. In other words, the underlying treatment effect (whether beneficial or harmful) that all studies are estimating is assumed to be the same for all studies. Any variability among study results is assumed to be random and is ignored in producing a summary estimate of the effect.^{91,92}

In any meta-analysis, however, 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. Methods for detecting and describing heterogeneity are described in detail by Hardy and Thompson.⁹³ One might stratify the studies according to patient characteristics or study design features and investigate heterogeneity within and across strata. To the extent that the stratification explains the heterogeneity, the combined results would differ between strata and the heterogeneity within the strata would be reduced compared to the overall result. In addition to stratification, regression methods such as weighted least squares linear regression could be used to explore sources of heterogeneity.^{3,29,94–96} These might be important when various components of study design are correlated with each other, acting as potential confounders. Graphical methods for meta-analysis have also been proposed, that focus on issues related to heterogeneity.⁹⁷

There are also 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 only studies with similar designs, protocols, and patient populations. Hidden or unmeasured sources of among-study variability of results are taken into account by these random effect models through the incorporation of such variability into the weighting scheme when computing a weighted average summary estimate. The practical consequence of the random effect models is to produce wider confidence intervals than would otherwise be produced by the traditional methods.^{91,92} This approach is particularly useful when there is heterogeneity among study results, and exploratory analyses have failed to uncover any known sources of observed heterogeneity. However, random effect models should not be viewed as a panacea for unexplained heterogeneity. One danger is that a summary measure of heterogeneous studies may not really apply to any particular study population or study design, i.e., they lose information by averaging over potentially important study and population characteristics.³⁸

A second practical effect of random effect models, which is only apparent from examining the mathematics involved, is that they tend to assign relatively higher weights to small studies than the traditional methods would assign.⁹¹ This equalization of weights may have unwanted consequences in some circumstances (see section on “Publication bias,” below), and can lead to counterintuitive results, with very small studies making contributions to the summary equal to those of very large studies. A thorough discussion of the interpretation and application of fixed versus random effect models is presented by Hedges and Vevea.⁹⁸

Bayesian statistical methods are also being proposed with increasing frequency in the statistical literature.^{99–104} These include the so-called “confidence profile” method developed by Eddy and colleagues.^{105,106} These methods can incorporate into the analysis the investigator’s prior beliefs

about the size of an effect or about the factors biasing the observed effects. When the investigator has no prior beliefs about the effect, the results of the observed studies are sometimes used to estimate the components of the “prior” distribution. Thus, the final answers reflect the observed data very closely. In practice, when the investigator does not specify prior beliefs, the summary results are similar to those from standard methods, especially the random effect models described above.

Another approach¹⁰⁷ is based on summing the statistical evidence provided by each study for each value of the effect measure (e.g., the odds ratio). The value of the odds ratio with the maximum evidence (the maximum likelihood estimate) has usually proved the same as the pooled estimate produced by other methods, in the situations where this method has been used.¹⁰⁷ By providing a mathematical and graphical picture of what is occurring at other values of the effect estimate, this method also provides information on the contribution of each study to the total.

An important word of caution is that statistical tests of heterogeneity, i.e., formal statistical tests of the variability among the studies, suffer from a notorious lack of statistical power.¹⁰⁸ Thus, a finding of significant heterogeneity may safely be interpreted as meaning the studies are not all estimating the same parameter. A lack of statistical significance, however, may not mean that heterogeneity is not important in a data set or that sources of variability should not be explored.

Formulate Conclusions and Recommendations

As with all research, the conclusions of a meta-analysis should be clearly summarized, with appropriate interpretation of the strengths and weaknesses of the meta-analysis. Authors should clearly state how generalizable the result is, how definitive it is, and should outline the areas that need future research. Any hypotheses generated by the meta-analysis should be stated as such, and not as conclusions.

APPROACHES TO SELECTED METHODOLOGIC PROBLEMS IN META-ANALYSIS

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.^{38,109} Conversely, diversity of designs and results may provide an opportunity to understand the factors that modify the effectiveness (or toxicity) of a drug. In fact, it has been argued that because of the potential for bias in observational epidemiologic studies, exploring heterogeneity should be the main point of meta-analyses of such studies, rather than producing a single summary measure.^{6,33,38,110}

In addition to considering the clinical and methodologic differences among studies, prior to the actual combination of results, it is essential to evaluate the extent to which study results are combinable from a strictly statistical viewpoint. This would usually involve some kind of statistical test of the variability (heterogeneity) of results among studies. If the variability of odds ratios (on the natural logarithm scale), for example, is greater than that which could be attributed to sampling variability alone, one should question the wisdom and meaning of a combined result. Efforts should subsequently focus on exploring the reasons for the variability of results, rather than proceeding to combine them.

The generality of the question posed will clearly influence the generalizability of the result, but also may affect whether the primary studies involved are viewed as combinable or not. Because the set of available studies is likely to be heterogeneous with respect to design features, the choice of a more general question may be preferred to a very specific one. For example, Dickersin and Berlin¹³ suggest that a more general question might be “Is taking aspirin associated with decreased mortality

in patients who have already had a myocardial infarction?" rather than "Is administration of 325 milligrams of aspirin per day, started within seven days of a first documented myocardial infarction and taken for at least six months, in the absence of other preventive treatments, in patients followed for at least one year, associated with a decrease in subsequent cardiovascular mortality?". If addressing the second question, the meta-analyst may quickly find him or herself with only one available study. Diversity of study designs, on the other hand, can provide a more generalizable result than restriction to a very narrowly defined group of studies. Issues of study design, such as dose or duration of therapy, and how study design relates to study results, could be addressed through a series of exploratory analyses.¹³

As an example of the type of analysis that could be used to investigate study design issues, Berlin and colleagues,³ in a methodologic paper, examine data originally presented by Romieu and colleagues²⁰ on the relationship between duration of oral contraceptive use and risk of breast cancer. They show, using regression methods and stratified analyses, that case-control studies involving mostly premenopausal women show a marked and quite homogeneous increase in risk with increasing duration of use of oral contraceptives. Studies including postmenopausal cases, conversely, show no such increase in risk, on average. Furthermore, the average effect is difficult to interpret, since the results of the postmenopausal studies are quite heterogeneous, with some studies showing an increase in risk with increasing duration of oral contraceptive use and others showing a decrease. One technical difficulty with meta-analysis of group-level data is also illustrated by this example. Although some studies present age-specific results, or are restricted to certain age groups, many studies do not present subgroup analyses that are of interest in a meta-analysis. Thus, for example, some studies that include postmenopausal women also include premenopausal women, and the data within such studies may not be presented by menopausal status.

In the situation just described, there were strong biological reasons for performing separate analyses by menopausal status. 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. Morgenstern *et al.*²⁸ 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. A practical technique for determining the potential for publication bias is the "funnel plot," first proposed by Light and Pillemer.¹¹¹ The method involves plotting the effect size (e.g., the risk difference) against a measure of study size, such as the sample size, or the inverse of the variance of the individual effect sizes. If there is no publication bias, the points should produce a kind of funnel shape, with a scatter of points centered around the true value of the effect size, and with the degree of scatter narrowing as the variances decrease. If publication bias is a problem, the funnel would look as though a bite had been taken out, with very few (if any) points around the point indicating no effect (e.g., odds ratio of 1.0) for studies with large variances. This method requires a sufficient number of studies to permit the visualization of a funnel shape to the data. If the funnel plot does indicate the existence of publication bias, then one

or more of the correction methods described below should be considered. In the presence of publication bias, the responsible meta-analyst should also evaluate the ethics of presenting a summary result that is likely to represent an overestimate of the effect in question.

Two examples of funnel plots are given in Figures 38.1 and 38.2. These plots represent studies of psychoeducational programs for surgical patients.^{111,112} In the first plot, only the published studies are represented. The funnel appears to have a “bite” taken out of it where the small studies showing no effect of these programs should be. In the second plot, the unpublished studies, including doctoral dissertations, are included, and the former “bite” is now filled with these unpublished studies.

Several mathematical approaches to the problem of publication bias have also been proposed. An early method, first described by Rosenthal,¹¹³ is the calculation of a “fail-safe N ” when the result of the meta-analysis is a statistically significant rejection of the null hypothesis. This method, in a kind of sensitivity analysis, uses the Z -statistics from the individual studies included in a meta-analysis to calculate the number of *unpublished*

studies with a Z -statistic of exactly 0 that would be required to exist, in order for the combined Z -score (published + unpublished studies) to become nonsignificant. Because this method focuses only on Z -statistics, and ignores the estimation of effects (e.g., odds ratios), it is of limited utility. That is, the fail-safe N approach focuses only on the statistical significance of the combined result and does not help provide an overall estimate of the effect that is “adjusted” for publication bias.

A number of related methods to deal with potential unpublished studies have been developed in recent years. These include other methods for estimating the number of unpublished studies,^{114,115} formal methods to test for the presence of publication bias,^{116–118} and methods to adjust summary estimates to account for unpublished studies,^{114,119–122} but several of those methods make some fairly strong assumptions about the specific mechanism producing the publication bias. The available data are used to estimate both the average effect size for the comparison of groups and the probabilities of publication. (Note that only published data are used to estimate the probability of publication, strengthening the dependence of such methods on the underlying

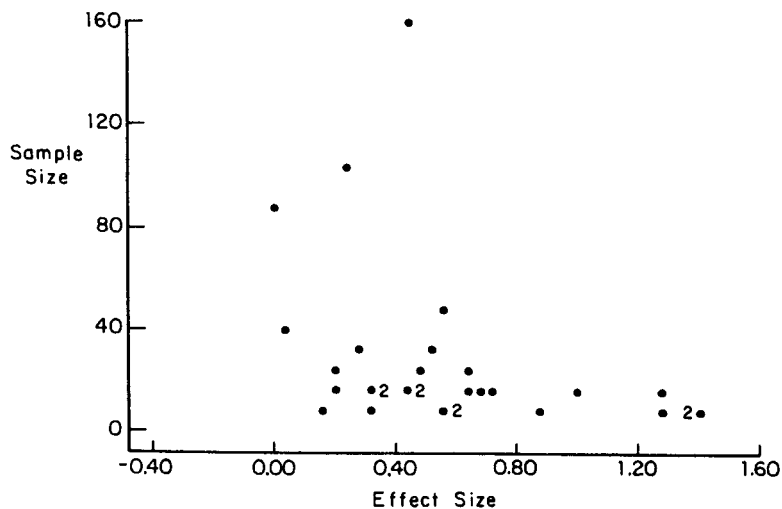


Figure 38.1. Funnel plot for published studies only: analysis of data from the review by Devine and Cooks¹¹² of psychoeducational programs for surgical patients. Reprinted by permission of the publishers from *Summing Up: the Science of Reviewing Research* by Richard J. Light and David B. Pillemer, Cambridge, MA: Harvard University Press, copyright 1984 by the President and Fellows of Harvard College.

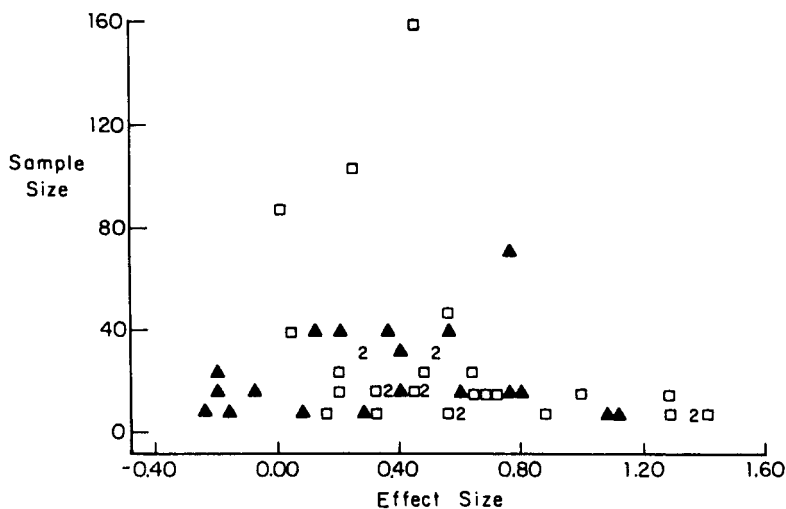


Figure 38.2. Funnel plot for published (open boxes) and unpublished (closed triangles) studies combined: analysis of data from the review by Devine and Cooks¹¹² of psychoeducational programs for surgical patients. Reprinted by permission of the publishers from *Summing Up: the Science of Reviewing Research* by Richard J. Light and David B. Pillemer, Cambridge, MA: Harvard University Press, copyright 1984 by the President and Fellows of Harvard College.

assumptions regarding the nature of the relationship between study findings and publication probabilities.)

An additional methodologic caution relates to the use of random effect models for combining results. When the results of the studies being analyzed are heterogeneous and a random effect model is being used to combine those results, one of the properties of the model, described above, is to assign relatively higher weights to small studies than would otherwise be assigned by more traditional methods of combining data. If publication bias is a problem in a particular data set, one consequence implied by the funnel plot is that small studies would tend to show larger effects than large studies. Thus, if publication bias is present, one of the reasons for heterogeneity of study results is that the small studies show systematically larger effects than the large studies. The assignment of higher relative weights to the small studies could, when publication bias is present, lead to a biased summary result.³⁸ In fact, this appears to be exactly the situation presented by Poole and Greenland in an examination of studies of water chlorination and cancer.³⁵ Random effect summary estimates of the relative risk

for various cancers were larger than corresponding fixed effect summaries. This was apparently due to the assignment of higher relative weights to small studies which, in this case, showed relatively larger effects, that may not be representative of the findings of all small studies.

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.^{47, 123–129} Going one step further, several prospectively planned meta-analyses are either being planned or have been conducted.^{130–133}

CASE STUDIES OF APPLICATIONS OF META-ANALYSIS

Investigation of Adverse Effects

As mentioned earlier, the investigation of adverse or unwanted effects of existing therapies is an important application of meta-analysis. As discussed in Chapter 3, 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.¹³⁴

The assessment of the excess risk of gastrointestinal side-effects associated with nonsteroidal anti-inflammatory drugs (NSAIDs) provides an excellent example of a situation in which meta-analysis has been helpful. Four different meta-analytic approaches to this problem will be reviewed here.

Chalmers and colleagues examined data from randomized trials of NSAIDs.³⁹ 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, excluding trials involving topical usage of NSAIDs, those that examined pharmacological endpoints only, studies of newborns, trials involving fewer than four days of treatment, trials in which patients were taking NSAIDs within the three days before randomization, and trials of drugs for dysmenorrhea (because of the short duration of the drug regimen and the confounding gastrointestinal symptoms from the dysmenorrhea). The meta-analysis was 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.

Data were extracted for the following endpoints: nausea, indigestion or dyspepsia, gross gastrointestinal bleeding, suspected ulcer only, proven ulcer, gastric side-effects, and unspecified gastric side-effects. Factors that might be related to the incidence of these endpoints were also extracted from the studies: disease under study, drug ingested, dose and duration of drug, age of patients, sex of patients, and date of publication. The data were analyzed by crude pooling (i.e., ignoring the stratification by study and simply collapsing over studies), by unweighted averaging of the within-study risk differences, and by a weighted average of the risk differences. Additional analyses were performed to determine whether any factors seemed to be associated with a study showing a harmful effect of NSAIDs.

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 nonaspirin 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 ten 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 followup for these ten 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 d) than for studies showing no effect of NSAIDs (geometric mean = 25 d) 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,⁸⁷ 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,⁹⁰ whose results suggest that the one-step method would have been the most appropriate for these studies with frequent occurrence of zero cells.

Another meta-analytic approach to the problem of side-effects of NSAIDs was used by Gabriel and colleagues,¹³⁵ who examined the results of 16 non-experimental studies (nine case-control and seven cohort) of serious gastrointestinal complications related to use of NSAIDs. The studies had to have a comparison group and provide an estimate of risk for serious gastrointestinal complications (defined as bleeding, perforation, or other adverse gastrointestinal events resulting in hospitalization or death) in NSAID users compared with nonusers, regardless of underlying disease. They excluded studies if the primary goal was to assess effectiveness.

The odds ratio found by these authors for gastrointestinal bleeding, based on nine studies reporting this endpoint, was 2.39 (CI 2.11, 2.70). The authors performed separate analyses for case-

control and cohort studies. Although these separate summaries are only reported graphically and the exact values are difficult to read, the summary odds ratio from the cohort studies is clearly closer to unity than the result from the case-control studies. These authors also found that the size of the odds ratio was related to the duration of NSAID use. Interestingly, though, the highest odds ratios were obtained from studies in which the duration of NSAID consumption was less than 1 month. (Note: Gabriel *et al.* only presented this finding without adjustment for study design, case-control versus cohort. Although they performed a multiple regression to examine interstudy heterogeneity, the findings of that model with respect to the individual potential sources of heterogeneity were not presented. It is possible that the studies with under one month of NSAID use were also predominantly case-control studies, but that cannot be determined from their paper, leaving the underlying source of heterogeneity somewhat ambiguous.)

The consistency of results for gastrointestinal bleeding between the two meta-analyses is of interest and lends some support to a causal association. Several points are important in considering the above results. In a cohort study of gastrointestinal bleeding and NSAIDs, Carson and colleagues¹³⁶ found a quadratic duration-response relation. They argue that this is compatible with an increasing risk with increasing duration of NSAID use, as suggested by Chalmers *et al.*,³⁹ until many of those patients who would develop gastrointestinal bleeding from NSAIDs were removed from the cohort and then the risk declined. This reasoning may explain the apparently anomalous finding by Gabriel *et al.*¹³⁵ of highest odds ratios for studies with less than one month of NSAID use.

Bollini and colleagues¹³⁷ also examined epidemiologic studies that investigated the association between NSAIDs and severe upper gastrointestinal tract disease, including hematemesis, melena, peptic ulcer, ulcer perforation, and death attributable to these outcomes. The studies had to compare groups exposed and unexposed to NSAIDs. Of the 34 studies they examined, seven were cohort, eight case-control with community

controls, and 19 case-control with hospital controls. The type of study design was associated with varying estimates of the relative risk. Case-control studies with hospital controls had the highest average relative risk (4.4 [3.3–6.0]) and the cohort studies had the lowest (2.0 [1.2–3.2]). They found that studies with satisfactory methods yielded on average a lower relative risk (2.6 [1.8–3.9]) as compared with studies whose methods were unsatisfactory (4.2 [3.1–5.6]).

In perhaps the most comprehensive and clinically useful of the systematic reviews in the area of NSAIDs side effects, Henry and colleagues¹³⁸ addressed the issue of *comparative* relative risks of serious gastrointestinal complications with individual NSAIDs. 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.

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 (95% CI 1.0, 2.5) for fenopfen, to 9.2 (95% 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.*¹³⁹ and Carson and Strom¹⁴⁰ 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 meta-analysis of Gabriel *et al.* Cohort studies based on claims data, such as that conducted by Carson and colleagues¹³⁶ described above, 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 NSAID, 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. These include the following.

1. Two meta-analyses have been published on the effects of prophylactic lidocaine in acute myocardial infarction.¹⁴¹⁻¹⁴³ These studies showed that, although lidocaine effectively prevented ventricular fibrillation,¹⁴³ there seemed to be an excess in mortality among those patients randomly allocated to lidocaine compared with those allocated to placebo. In a related paper, using meta-analytic regression methods, Antman and Berlin¹⁴⁴ calculate that, given the low baseline incidence of ventricular fibrillation in the current coronary care unit environment, 400 patients would require treatment with lidocaine to prevent a single episode of ventricular fibrillation. Considering this estimate in addition to the possibly increased risk of mortality, the authors suggest that prophylactic lidocaine should not be given routinely.
2. In a meta-analysis of randomized trials regarding the efficacy and safety of quinidine therapy for the maintenance of sinus rhythm after cardioversion,¹⁴⁵ the authors show that quinidine is, indeed, effective at maintaining sinus rhythm, but that mortality seems to be elevated in quinidine patients (mortality odds ratio = 2.98, 95% CI 1.1, 8.3, based on 12 deaths among patients randomized to quinidine versus only three among patients randomized to placebo). In a subsequent paper, the authors

examine the relationship between study design and outcome.¹⁴⁶ The nonrandomized studies tended to show less benefit of quinidine with respect to maintenance of sinus rhythm compared with the randomized studies. The odds ratios for mortality also varied according to study design, although there were few deaths overall: OR = 3.5 (1.0, 12.4) for randomized studies; OR = 9.9 (0.8, 123.2) for nonrandomized studies. The overall mortality risk was 2.0% (34/1709) for quinidine-treated patients and 0.6% (4/681) for all control patients.

3. A third example is the meta-analysis of nonexperimental studies of oral contraceptives and the risk of breast cancer discussed above.²⁰ An association between increasing odds ratio and increasing duration of oral contraceptive use was found in case-control studies in which the cases were mostly premenopausal (defined as age limit of 45 years old or less). In a subsequent methodologic paper, Berlin and colleagues³ use these same data to show that the magnitude of the odds ratio depends not only on the menopausal status of the cases but on the calendar years during which cases were accrued, presumably because of the changing formulations of oral contraceptives.

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 performed a meta-analysis of the published data from the randomized trials of induction therapy¹⁴⁷ 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 examined the effect of induction therapy beyond two years and in subgroups of patients with risk factors for early allograft failure.¹⁴⁸ The subgroup analyses are examined in the next section. The five studies included in the individual patient analyses gave 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 it was no longer significant at five years.

Differential Effects Among Subgroups of Patients

In the analysis of individual patient data by Szczech and colleagues, 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 level (PRA), 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.¹⁴⁸

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.

In another example of a meta-analysis addressing subgroup issues, Midgette and colleagues¹⁴⁹ examined the use of intravenous streptokinase in patients with suspected acute myocardial infarction. The authors found a summary relative risk of 0.72 (0.65, 0.79), favoring streptokinase, in patients with suspected acute anterior myocardial infarction, and a summary risk ratio of 0.87 (0.76, 1.01) in those with a suspected inferior infarction. The authors conclude that there is a protective effect of intravenous streptokinase in anterior infarction, but that the protective effect in inferior infarction is smaller and less certain.

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.*¹⁵⁰ 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 β -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 β -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 β -blockers. The authors drew no firm conclusions from the subgroup analyses other than to suggest directions for future research.

As another example, for several decades, heparin has been used as the primary antithrombotic drug for the initial treatment of venous thromboembolism. There has been some controversy over the optimal mode of administration of heparin: intermittent intravenous, continuous intravenous, or subcutaneous injection. While continuous infusion has been shown to be safer than intermittent injection, and equally effective, continuous infusion has disadvantages, such as the need for hospitalization for most patients, possibly prolonged immobilization, enhanced risk for sepsis related to the infusion cannula, and possible increased cost.¹⁵¹ Hommes and colleagues performed a meta-analysis of randomized trials of intravenous versus subcutaneous heparin administration in the initial treatment of deep vein thrombosis.¹⁵¹ They found eight studies meeting their inclusion criteria. The overall summary relative risk for efficacy was 0.62 (0.39, 0.98), indicating a benefit from the use of subcutaneous compared with intravenous administration. The analysis of safety (i.e., risk of major hemorrhage) also showed a modest benefit of subcutaneous injection (relative risk = 0.79, CI 0.42, 1.46). In the analysis of efficacy, as in others described above, there was highly significant among study heterogeneity. The source of this heterogeneity was apparently a single study showing a significant

benefit from intravenous administration.¹⁵² The authors of the meta-analysis speculate that the particular study failed to achieve therapeutic levels of heparin in the subcutaneous group.

A somewhat different approach to subgroup analyses has emerged in recent years. This strategy views the risk of events in the control group of a trial (baseline risk) as a general indicator of severity of disease in the treated population. The relationship between treatment benefit and baseline risk can then be estimated, i.e., examining whether there is an interaction between treatment and baseline risk.^{101, 102, 153} A number of statistical issues arise in such analyses, including the inherent association induced by regressing treatment benefit (e.g., the log relative risk, which is calculated using the baseline risk) on baseline risk, and the fact that the baseline risk, except in very large studies, is affected by sampling variability. These approaches all use group level (published) data, examining whether the risk in the *population* is associated with the magnitude of treatment benefit. This may not necessarily yield the same result as looking at the interaction on an *individual patient* basis between treatment and *individual level* estimates of risk. It is also clearly different clinically from examining specific patient characteristics, as opposed to calculating estimates of risk that depend on several characteristics. Information may be lost by using multivariable risk estimates, as opposed to potentially biologically specific patient characteristics.

Saving Time and Money If You Believe a Meta-Analysis

One of the potential benefits of meta-analysis is the potential to shorten the time between a medical research finding and the clinical implementation of a new therapy. This is a concern not only for the development of new drugs, but for the exploration of new indications for existing therapies. A group at Harvard University 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.^{40, 154} Antman *et al.*⁴⁰ 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.

As an example, Antman and colleagues analyzed data from 17 trials of β -blockers for the prevention of death in the years following a myocardial infarction.⁴⁰ On the left hand side of Figure 38.3, reproduced from their paper, the data are presented as a traditional meta-analysis, with individual study results presented along with the

summary odds ratio arbitrarily estimated after 17 trials had been completed. On the right hand side of Figure 38.3, the same data are presented as a cumulative meta-analysis, with an updated summary estimate calculated after the completion of each new trial. The cumulative meta-analysis clearly shows that the updated pooled estimate became statistically significant in 1977 and has remained so ever since.

The process of cumulative meta-analysis was applied by these authors to eight therapies for acute myocardial infarction. In five of the six instances in which the cumulative meta-analyses

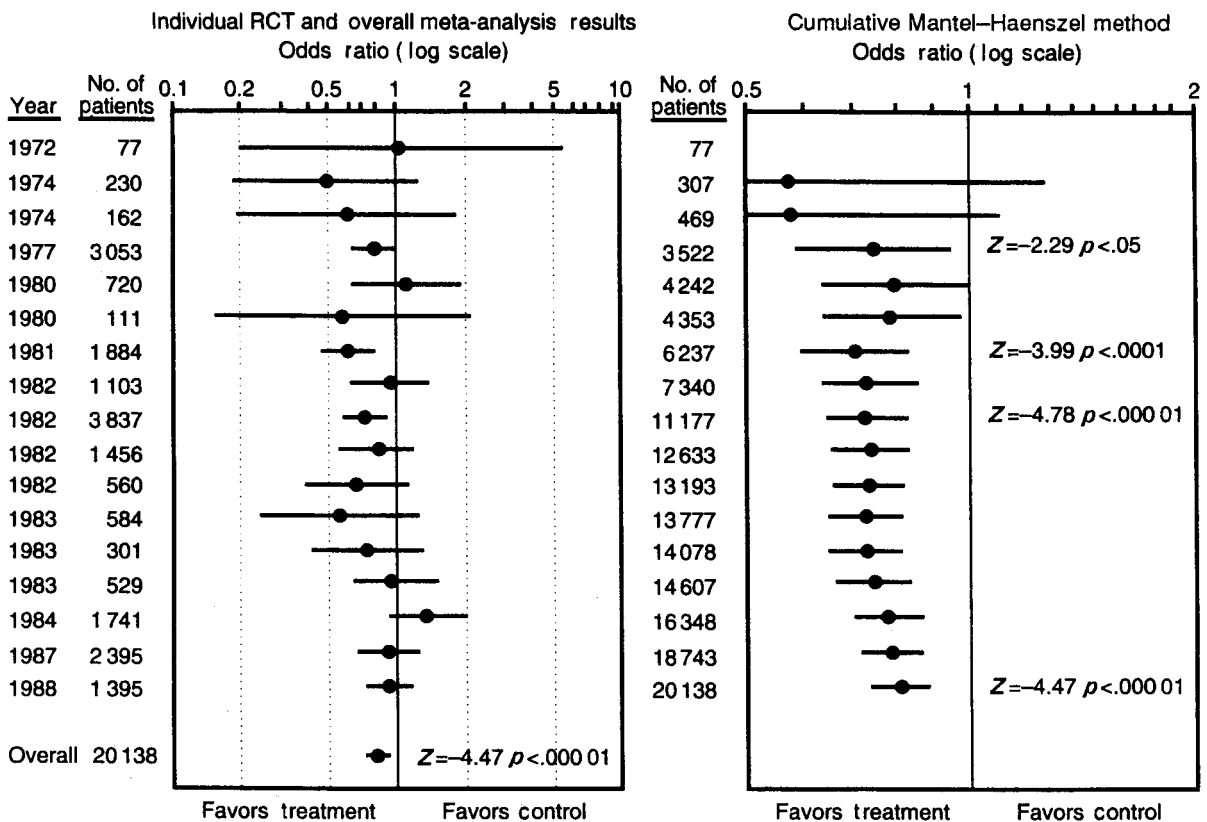


Figure 38.3. Results of 17 randomized control trials of the effects of oral β -blockers for secondary prevention of mortality in patients surviving a myocardial infarction presented as two types of meta-analysis. On the left is the traditional one, revealing many trials with nonsignificant results but a highly significant estimate of the pooled results on the bottom of the panel. On the right, the same data are presented as cumulative meta-analyses, illustrating that the updated pooled estimate became statistically significant in 1977 and has remained so up to the present. Note that the scale is changed on the right-hand graph to improve clarity of the confidence intervals. Reprinted with permission from Antman *et al.*, *Journal of the American Medical Association*; July 8, 1992; volume 268; pages 240–248. Copyright 1992, American Medical Association.

revealed the therapies to be of statistically significant benefit in reducing in-hospital mortality, it was several years before experts recommended the therapy with any consistency. An important example was thrombolytic therapy, which did not begin to be recommended by more than half the experts, even for specific indications, until 13 years after the cumulative meta-analysis would have shown therapy to be effective. Six years passed between the publication, in a major journal,^{40,155} of the first meta-analysis showing an impressive reduction in mortality by thrombolytic therapy and the year in which the majority of the experts whose opinions were studied by the authors recommended it for routine or specific use. In 1985, a 20% reduction in mortality was established at the $p < 0.001$ level (OR = 0.78; CI 0.69, 0.90). A total of 14 reviews after 1985 did not mention the treatment or felt that it was still experimental. The authors concluded that identifying and interpreting the therapeutic trials in a given field is extremely difficult, so that clinical experts need access to better databases and new statistical techniques (such as cumulative meta-analysis).

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.^{100,156} At the least, one might wish to consider using a more stringent criterion for statistical significance than the traditional $p < 0.05$ cutoff. Another consideration is that estimates of treatment effect may not be stable over time, perhaps due to changing clinical environments. In the β -blocker example, there is an apparent "drift" of the effect estimate back toward the null in more recent years, i.e., treatment appears to be less effective in the most recent studies. Thus, it may be important to re-evaluate therapies as other treatment strategies evolve for the same conditions.

A final caution with regard to interpreting cumulative meta-analyses relates to the continuing need for well designed randomized controlled trials. New indications for existing therapies, for example,

are often suggested by nonexperimental studies, including cohort and case-control studies and nonrandomized phase 2 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 (see Ziegler *et al.*¹⁵⁷ for a review) 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.¹⁵⁸⁻¹⁶⁰

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 meta-regression techniques on what amounts to ecological data. As a trivial hypothetical example, suppose one wanted to perform separate analyses of the effect of treatment X in men and women. Among the existing randomized trials, six exclude women and four do not, but the four also include men. Ideally,

one would like to obtain information on the effect in men alone from all of the ten trials, since all include men. Similarly, one could obtain a separate estimate of the effect of treatment in women from the four studies including women. It is possible to use the group level data only to show, for example, that studies that include women tend to show different effects than studies excluding women, but one cannot perform a separate analysis of women. One might alternatively regress the treatment effect measure (e.g., log relative risk) against the percent male (or female), but that is still less satisfactory than obtaining patient level data. For further practical discussions of the use of individual patient data, see Stewart and Clarke¹⁶¹ and Stewart and Parmar.¹⁶² Mechanisms for the sharing of person level data need to be promoted.

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? Suppose ten studies all show a 20% (or thereabouts) reduction in mortality in patients treated with drug X compared to placebo, but none of the individual studies shows a statistically significant effect. If the combined analysis shows a highly statistically significant 20% reduction in mortality due to treatment X, and 20% is considered clinically significant, should the combined analysis be sufficient evidence for the acceptance of drug X as beneficial? What would an additional, large clinical trial contribute?

In this context, it is worth noting that several empirical studies have examined discrepancies between large trials and meta-analyses of the same therapies.^{163–168} 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. For example, for BCG vaccine, a huge trial using passive followup, and therefore missing (by the authors'

own admission) at least 50% of all cases of tuberculosis, failed to show the protective effect found in a number of other studies with more complete followup.^{169, 170}

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. (Consider the slow pace, demonstrated by Antman *et al.*,⁴⁰ with which new therapies are accepted even when the evidence is provided by randomized trials.) Even if a meta-analysis of, for example, oral contraceptive use and breast cancer risk were to show a convincing, consistent duration–response relationship, the issue of what to do with that information is complex. If the relative risk for 10 years of use is 1.5, is that sufficient to warrant removal of oral contraceptives from the market? Would 2.5 be a sufficiently high relative risk? What other factors, e.g., family history, etc., need to be considered when prescribing oral contraceptives? While these are clearly more general issues, not restricted to the interpretation of meta-analyses, the additional precision provided by meta-analyses makes their interpretation all the more difficult.

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.^{130–133} 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 for a more complete discussion of this issue).³³

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 pharmacoepidemiology research.

REFERENCES

1. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–88.
2. Last JM. *A Dictionary of Epidemiology*. New York: Oxford University Press, 1988.
3. Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. *Epidemiology* 1993; **4**: 218–28.
4. Feinstein AR. Para-analysis, faute de mieux, and the perils of riding on a data barge. *J Clin Epidemiol* 1989; **42**: 929–35.
5. Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *J Clin Epidemiol* 1997; **50**: 223–29.
6. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *Br Med J* 1998; **316**: 140–4.
7. Thacker SB. Meta-analysis. A quantitative approach to research integration. *J Am Med Assoc* 1988; **259**: 1685–9.
8. Buffler PA. The evaluation of negative epidemiologic studies: the importance of all available evidence in risk characterization. *Regul Toxicol Pharmacol* 1989; **9**: 34–43.
9. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *New Engl J Med* 1987; **316**: 450–5.
10. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987; **107**: 224–33.
11. Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. *Am J Epidemiol* 1995; **142**: 371–82.
12. Berlin JA. Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *Am J Epidemiol* 1995; **142**: 383–7.
13. Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. *Epidemiol Rev* 1992; **14**: 154–76.
14. Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987; **106**: 485–8.
15. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988; **138**: 697–703.
16. Chalmers I. Improving the quality and dissemination of clinical research. In: Lock S, ed., *The Future of Medical Journals: in Commemoration of 150 Years of the British Medical Journal*. London: British Medical Society, 1991; 127–46.
17. Kass EH. Reviewing reviews. In: Coping with the biomedical literature: a primer for the scientist and the clinician. Warren KS, ed. New York: Praeger, 1981. 71–91.
18. Huth EJ. Needed: review articles with more scientific rigor. *Ann Intern Med* 1987; **106**: 470–1.
19. Canner PL. An overview of six clinical trials of aspirin in coronary heart disease. *Stat Med* 1987; **6**: 255–67.
20. Romieu I, Berlin JA, Colditz G. Oral contraceptives and breast cancer. Review and meta-analysis. *Cancer* 1990; **66**: 2253–63.
21. Longnecker MP, Berlin JA, Orza MJ, Chalmers TC. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *J Am Med Assoc* 1988; **260**: 652–6.
22. Longnecker MP, Orza MJ, Adams ME, Vioque J, Chalmers TC. A meta-analysis of alcoholic beverage consumption in relation to risk of colorectal cancer. *Cancer Causes Control* 1990; **1**: 59–68.
23. Wong O, Raabe GK. Critical review of cancer epidemiology in petroleum industry employees, with a quantitative meta-analysis by cancer site. *Am J Ind Med* 1989; **15**: 283–310.
24. Frumkin H, Berlin J. Asbestos exposure and gastrointestinal malignancy: review and meta-analysis. *Am J Ind Med* 1988; **14**: 79–95 [published erratum appears in *Am J Ind Med* 1988; **14**: 493].

25. Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol* 1990; **132**: 612–28.
26. Gray A, Berlin JA, McKinlay JB, Longcope C. An examination of research design effects on the association of testosterone and male aging: results of a meta-analysis. *J Clin Epidemiol* 1991; **44**: 671–84.
27. Booth-Kewley S, Friedman HS. Psychological predictors of heart disease: a quantitative review. *Psychol Bull* 1987; **101**: 343–62.
28. Morgenstern H, Glazer WM, Niedzwiecki D, Nourjah P. The impact of neuroleptic medication on tardive dyskinesia: a meta-analysis of published studies. *Am J Public Health* 1987; **77**: 717–24.
29. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 1987; **9**: 1–30.
30. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose–response data, with applications to meta-analysis. *Am J Epidemiol* 1992; **135**: 1301–9.
31. Greenland S, Salvan A. Bias in the one-step method for pooling study results. *Stat Med* 1990; **9**: 247–52.
32. Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J Clin Epidemiol* 1991; **44**: 127–39.
33. Berlin JA, Colditz GA. The role of meta-analysis in the regulatory process for foods, drugs, and devices. *J Am Med Assoc* 1999; **281**: 830–4.
34. O'Neill RT, Anello C. Does research synthesis have a place in drug regulatory policy? Synopsis of issues: assessment of efficacy and drug approval. *Clin Res Regul Affairs* 1996; **13**: 23–9.
35. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol* 1999; **150**: 469–75.
36. Jones DR. Meta-analysis of observational epidemiological studies: a review. *J R Soc Med* 1992; **85**: 165–8.
37. Einarson TR, Leeder JS, Koren G. A method for meta-analysis of epidemiological studies. *Drug Intell Clin Pharm* 1988; **22**: 813–24.
38. Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994; **140**: 290–6.
39. Chalmers TC, Berrier J, Hewitt P, Berlin J, Reitman D, Nagalingam R, Sacks H. Meta-analysis of randomized controlled trials as a method of estimating rare complications of non-steroidal anti-inflammatory drug therapy. *Aliment Pharmacol Ther* 1988; **2**: 9–26.
40. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *J Am Med Assoc* 1992; **268**: 240–8.
41. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989; **81**: 107–15.
42. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337**: 867–72.
43. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *J Am Med Assoc* 1992; **267**: 374–8.
44. Dickersin K. The existence of publication bias and risk factors for its occurrence. *J Am Med Assoc* 1990; **263**: 1385–9.
45. Chalmers I. Underreporting research is scientific misconduct. *J Am Med Assoc* 1990; **263**: 1405–8.
46. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *J Am Med Assoc* 1990; **263**: 1392–5.
47. Moher D, Berlin J. Improving the reporting of randomised controlled trials. In: Non-random reflections on health services research. Maynard A, Chalmers I, eds., London: BMJ, 1997. 250–71.
48. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993; Doc. No. 50.
49. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *Br Med J* 1997; **315**: 640–5.
50. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *J Am Med Assoc* 1998; **279**: 281–6.
51. Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 1989; **84**: 381–92.
52. Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. II: Replicate variability and comparison of studies that agree and disagree. *Stat Med* 1987; **6**: 733–44.
53. Early Breast Cancer Trialists' Collaborative Group. *Treatment of Early Breast Cancer. Vol I. Worldwide Evidence*. New York, NY: Oxford University Press, 1990.
54. Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989; **84**: 374–80.
55. Gotzsche PC. Reference bias in reports of drug trials. *Br Med J Clin Res Edn* 1987; **295**: 654–6.

56. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997; **350**: 326–9.
57. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, Liberati A, Linde K, Penna A. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; **347**: 363–6.
58. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986; **1**: 155–8.
59. Hemminki E. Study of information submitted by drug companies to licensing authorities. *Br Med J* 1980; **280**: 833–6.
60. Cho MK, Bero LA. The quality of drug studies published in symposium proceedings. *Ann Intern Med* 1996; **124**: 485–9.
61. Chalmers TC. Problems induced by meta-analyses. *Stat Med* 1991; **10**: 971–9.
62. Conn HO, Blitzer BL. Nonassociation of adrenocorticosteroid therapy and peptic ulcer. *New Engl J Med* 1976; **294**: 473–9.
63. Messer J, Reitman D, Sacks HS, Smith H Jr, Chalmers TC. Association of adrenocorticosteroid therapy and peptic-ulcer disease. *New Engl J Med* 1983; **309**: 21–4.
64. Conn HO, Poynard T. Adrenocorticosteroid administration and peptic ulcer: a critical analysis. *J Chron Dis* 1985; **38**: 457–68.
65. Chalmers TC. Meta-analysis in clinical medicine. *Trans Am Clin Climatol Assoc* 1987; **99**: 144–50.
66. DerSimonian R. Parenteral nutrition with branched-chain amino acids in hepatic encephalopathy: meta analysis. *Hepatology* 1990; **11**: 1083–4.
67. Eriksson LS, Conn HO. Branched-chain amino acids in hepatic encephalopathy. *Gastroenterology* 1990; **99**: 604–7 [published erratum appears in *Gastroenterology* 1990; **99**: 1547].
68. Naylor CD, O'Rourke K, Detsky AS, Baker JP. Parenteral nutrition with branched-chain amino acids in hepatic encephalopathy. A meta-analysis. *Gastroenterology* 1989; **97**: 1033–42.
69. Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. *Control Clin Trials* 1985; **6**: 306–17.
70. Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature. A comparison of MEDLARS and manual methods. *Control Clin Trials* 1985; **6**: 271–9.
71. Bernstein F. The retrieval of randomized clinical trials in liver diseases from the medical literature: manual versus MEDLARS searches. *Control Clin Trials* 1988; **9**: 23–31.
72. Chalmers I, ed. Oxford database of perinatal trials. Oxford: Oxford University Press, 1988.
73. Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996; **347**: 1024–6.
74. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992; **45**: 255–65.
75. Berard A, Bravo G. Combining studies using effect sizes and quality scores: application to bone loss in postmenopausal women. *J Clin Epidemiol* 1998; **51**: 801–7.
76. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; **2**: 31–49.
77. Imperiale TF, McCullough AJ. Do corticosteroids reduce mortality from alcoholic hepatitis? A meta-analysis of the randomized trials. *Ann Intern Med* 1990; **113**: 299–307.
78. Prendiville W, Elbourne D, Chalmers I. The effects of routine oxytocic administration in the management of the third stage of labour: an overview of the evidence from controlled trials. *Br J Obstet Gynaecol* 1988; **95**: 3–16.
79. Mahon WA, Daniel EE. A method for the assessment of reports of drug trials. *Can Med Assoc J* 1964; **90**: 565–9.
80. Verhagen AP, de Vet HCW, de Bie RA, Kessels AGH, Boers M, Bouter LM, Knipschild PG. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998; **51**: 1235–41.
81. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care* 1996; **12**: 195–208.
82. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995; **16**: 62–73.
83. Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994; **140**: 300–1.
84. Juni P, Witschi A, Block R, Egger M. The hazards of scoring the quality of clinical trials: lessons for meta-analysis. *J Am Med Assoc* 1999; **282**: 1054–60.
85. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J Am Med Assoc* 1995; **273**: 408–12.

86. Berlin JA, Miles CG, Cirigliano MD, Conill AM, Goldmann DR, Horowitz DA, Jones F, Scott E, Hanchak NA, Williams SV. Does blinding of readers affect the results of metaanalyses? Results of a randomized trial. *Online J Curr Clin Trials* 1997; Doc. No. 205.
87. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. New York: Van Nostrand Reinhold, 1982.
88. Rothman KJ. Modern epidemiology. Boston, MA: Little, Brown, 1986.
89. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985; **27**: 335–71.
90. 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*.
91. Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989; **8**: 141–51.
92. Fleiss JL. The statistical basis of meta-analysis. *Stat Methods Med Res* 1993; **2**: 121–45.
93. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Stat Med* 1998; **17**: 841–56.
94. Berlin JA, Antman EM. Advantages and limitations of metaanalytic regressions of clinical trials data. *Online J Curr Clin Trials* 1994; Doc. No. 134.
95. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Stat Med* 1995; **14**: 395–411.
96. Vanhonerack WR. Meta-analysis and response surface extrapolation: a least squares approach. *Am Stat* 1996; **50**: 294–9.
97. Walker AM, Martin-Moreno JM, Artalejo FR. Odd man out: a graphical approach to meta-analysis. *Am J Public Health* 1988; **78**: 961–6.
98. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychological Methods* 1998; **3**: 486–504.
99. Carlin JB. Meta-analysis for 2×2 tables: a Bayesian approach. *Stat Med* 1992; **11**: 141–58.
100. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Stat Med* 1991; **10**: 1665–77.
101. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Stat Med* 1997; **16**: 2741–58.
102. McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. *Stat Med* 1996; **15**: 1713–28.
103. Larose DT, Dey DK. Grouped random effects models for Bayesian meta-analysis. *Stat Med* 1997; **16**: 1817–29.
104. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995; **14**: 2685–99.
105. Eddy DM, Hasselblad V, Shachter R. An introduction to a Bayesian method for meta-analysis: the confidence profile method. *Med Decis Making* 1990; **10**: 15–23.
106. Eddy DM, Hasselblad V, Shachter R. Meta-analysis by the confidence profile method: the statistical synthesis of evidence. Boston, MA: Academic, 1990.
107. Goodman SN. Meta-analysis and evidence. *Control Clin Trials* 1989; **10**: 188–204. [published erratum appears in *Control Clin Trials* 1989; **10**: 435].
108. Fleiss JL. Analysis of data from multiclinic trials. *Control Clin Trials* 1986; **7**: 267–75.
109. Simon R. Overviews of randomized clinical trials. *Cancer Treat Rep* 1987; **71**: 3–5.
110. Davey Smith G, Egger M, Phillips AN. Meta-analysis. Beyond the grand mean? *Br Med J* 1997; **315**: 1610–14.
111. Light RJ, Pillemer DB. Summing up: the science of reviewing research. Cambridge, MA: Harvard University Press, 1984.
112. Devine EC, Cook TD. Effects of psycho-educational interventions on length of hospital stay: a meta-analytic review of 34 studies. In: Light RJ, ed., *Evaluation Studies Review Annual*, Vol. 8. Beverly Hills, CA: Sage, 1983.
113. Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979; **86**: 638–41.
114. Iyengar S, Greenhouse JB. Selection models and the file-drawer problem. *Stat Sci* 1988; **3**: 109–17.
115. Gleser LJ, Olkin I. Models for estimating the number of unpublished studies. *Stat Med* 1996; **15**: 2493–507.
116. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997; **315**: 629–34.
117. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; **50**: 1088–101.
118. Dear KBG, Begg CB. An approach for assessing publication bias prior to performing a meta-analysis. *Stat Sci* 1992; **7**: 237–45.
119. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic, 1985.
120. Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* 1995; **60**: 419–35.
121. Hedges LV, Vevea JL. Estimating effect size under publication bias: small sample properties and robustness of a random effects selection model. *J Educ Behav Stat* 1996; **21**: 299–332.
122. Givens GH, Smith DD, Tweedie RL. Publication bias in meta-analysis: a Bayesian data-augmentation approach to account for issues exemplified in

- the passive smoking debate. *Stat Sci* 1997; **12**: 221–50.
123. Meinert CL. Toward prospective registration of clinical trials. *Control Clin Trials* 1988; **9**: 1–5.
 124. Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986; **4**: 1529–41.
 125. Dickersin K. Report from the panel on the Case for Registers of Clinical Trials at the Eighth Annual Meeting of the Society for Clinical Trials. *Control Clin Trials* 1988; **9**: 76–81.
 126. Dickersin K. Why register clinical trials?—revisited. *Control Clin Trials* 1992; **13**: 170–7.
 127. Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *Br Med J* 1996; **313**: 1390–3.
 128. Anonymous. Making clinical trialists register. *Lancet* 1991; **338**: 244–5.
 129. Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. *Br Med J* 1992; **305**: 786–8.
 130. Valsecchi MG, Masera G. A new challenge in clinical research in childhood ALL: the prospective meta-analysis strategy for intergroup collaboration. *Ann Oncol* 1996; **7**: 1005–8.
 131. Margitic SE, Morgan TM, Sager MA, Furberg CD. Lessons learned from a prospective meta-analysis. *J Am Geriatr Soc* 1995; **43**: 435–9.
 132. Simes RJ. Prospective meta-analysis of cholesterol-lowering studies: the Prospective Pravastatin Pooling (PPP) Project and the Cholesterol Treatment Trialists (CTT) Collaboration. *Am J Cardiol* 1995; **76**: 122C–126C.
 133. Whitehead A. A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. *Stat Med* 1997; **16**: 2901–13.
 134. Temple RJ. The regulatory evolution of the integrated safety summary. *Drug Information J* 1991; **25**: 485–92.
 135. Gabriel SE, Jaakkimainen L, Bombardier C. Risk for serious gastrointestinal complications related to use of nonsteroidal anti-inflammatory drugs. A meta-analysis. *Ann Intern Med* 1991; **115**: 787–96.
 136. 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.
 137. Bollini P, Garcia Rodriguez L, Perez Gutthann S, Walker AM. The impact of research quality and study design on epidemiologic estimates of the effect of nonsteroidal anti-inflammatory drugs on upper gastrointestinal tract disease. *Arch Intern Med* 1992; **152**: 1289–95.
 138. Henry D, Lim LL-Y, Garcia Rodriguez LA, Perez Gutthann S, Carson JL, Griffin M, Savage R, Logan R, Moride Y, Hawkey C, Hill S, Fries JT. Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta-analysis. *Br Med J* 1996; **312**: 1563–6.
 139. Taragin MI, Carson JL, Strom BL. Gastrointestinal side effects of the nonsteroidal anti-inflammatory drugs. *Dig Dis* 1990; **8**: 269–80.
 140. Carson JL, Strom BL. The gastrointestinal toxicity of the non-steroidal anti-inflammatory drugs. In: Side-Effects of Anti-Inflammatory Drugs 3. Rainford KD, Velo GP, eds. Boston, MA: Kluwer, 1992; 1–8.
 141. Hine LK, Laird NM, Hewitt P, Chalmers TC. Meta-analysis of empirical long-term antiarrhythmic therapy after myocardial infarction. *J Am Med Assoc* 1989; **262**: 3037–40.
 142. Hine LK, Laird N, Hewitt P, Chalmers TC. Meta-analytic evidence against prophylactic use of lidocaine in acute myocardial infarction. *Arch Intern Med* 1989; **149**: 2694–8.
 143. MacMahon S, Collins R, Peto R, Koster RW, Yusuf S. Effects of prophylactic lidocaine in suspected acute myocardial infarction. An overview of results from the randomized, controlled trials. *J Am Med Assoc* 1988; **260**: 1910–16.
 144. Antman EM, Berlin JA. Declining incidence of ventricular fibrillation in myocardial infarction. Implications for the prophylactic use of lidocaine. *Circulation* 1992; **86**: 764–73.
 145. Coplen SE, Antman EM, Berlin JA, Hewitt P, Chalmers TC. Efficacy and safety of quinidine therapy for maintenance of sinus rhythm after cardioversion. A meta-analysis of randomized control trials. *Circulation* 1990; **82**: 1106–16 [published erratum appears in *Circulation* 1991; **83**: 714].
 146. Reimold SC, Chalmers TC, Berlin JA, Antman EM. Assessment of the efficacy and safety of antiarrhythmic therapy for chronic atrial fibrillation: observations on the role of trial design and implications of drug-related mortality. *Am Heart J* 1992; **124**: 924–32.
 147. 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.
 148. 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.
 149. Midgette AS, O'Connor GT, Baron JA, Bell J. Effect of intravenous streptokinase on early mortality in patients with suspected acute myocardial infarction. A meta-analysis by anatomic location of infarction. *Ann Intern Med* 1990; **113**:

- 961–8 [published erratum appears in *Ann Intern Med* 1991; **114**: 522].
150. Andrews TC, Reimold SC, Berlin JA, Antman EM. Prevention of supraventricular arrhythmias after coronary artery bypass surgery. A meta-analysis of randomized control trials. *Circulation* 1991; **84**: 236–44.
151. Hommes DW, Bura A, Mazzolai L, Buller HR, ten Cate JW. Subcutaneous heparin compared with continuous intravenous heparin administration in the initial treatment of deep vein thrombosis. A meta-analysis. *Ann Intern Med* 1992; **116**: 279–84.
152. Hull RD, Raskob GE, Hirsh J, Jay RM, Leclerc JR, Geerts WH, Rosenbloom D, Sackett DL, Anderson C, Harrison L. Continuous intravenous heparin compared with intermittent subcutaneous heparin in the initial treatment of proximal-vein thrombosis. *New Engl J Med* 1986; **315**: 1109–14.
153. Cook RJ, Walter SD. A logistic model for trend in $2 \times 2 \times k$ tables with applications to meta-analyses. *Biometrics* 1997; **53**: 352–7.
154. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *New Engl J Med* 1992; **327**: 248–54.
155. Stampfer MJ, Goldhaber SZ, Yusuf S, Peto R, Hennekens CH. Effect of intravenous streptokinase on acute myocardial infarction: pooled results from randomized trials. *New Engl J Med* 1982; **307**: 1180–2.
156. Berkey CS, Mosteller F, Lau J, Antman EM. Uncertainty of the time of first significance in random effects cumulative meta-analysis. *Control Clin Trials* 1996; **17**: 357–71.
157. Ziegler RG, Mayne ST, Swanson CA. Nutrition and lung cancer. *Cancer Causes Control* 1996; **7**: 157–77.
158. The Alpha-Tocopherol/Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New Engl J Med* 1994; **330**: 1029–35.
159. Hennekens CH, Buring JE, Manson JE, Stampfer M, Rosner B, Cook NR, Belanger C, LaMotte F, Gaziano JM, Ridker PM, Willett W, Peto R. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *New Engl J Med* 1996; **334**: 1145–9.
160. Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL, Valanis B, Williams JH, Barnhart S, Hammar S. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *New Engl J Med* 1996; **334**: 1150–5.
161. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. *Stat Med* 1995; **14**: 2057–79.
162. Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993; **341**: 418–22.
163. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New Engl J Med* 1997; **337**: 536–42.
164. Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. *J Am Med Assoc* 1998; **279**: 1089–93.
165. Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, Lau J. Large trials vs meta-analysis of smaller trials: how do their results compare? *J Am Med Assoc* 1996; **276**: 1332–8.
166. Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995; **48**: 23–40.
167. Villar J, Piaggio G, Carroli G, Donner A. Factors affecting the comparability of meta-analyses and largest trials results in perinatology. *J Clin Epidemiol* 1997; **50**: 997–1002.
168. Borzak S, Ridker PM. Discordance between meta-analyses and large-scale randomized, controlled trials. Examples from the management of acute myocardial infarction. *Ann Intern Med* 1995; **123**: 873–7.
169. Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, Mosteller F. The efficacy of bacillus Calmette–Guerin vaccination in the prevention of tuberculosis: meta-analysis of the published literature. *J Am Med Assoc* 1994; **271**: 698–702.
170. Colditz GA, Berkey CS, Mosteller F, Brewer TF, Wilson ME, Burdick E, Fineberg HV. The efficacy of bacillus Calmette–Guerin vaccination of newborns and infants in the prevention of tuberculosis: meta-analyses of the published literature. *Pediatrics* 1995; **96**: 29–35.

39

Validity of Pharmacoepidemiology Drug and Diagnosis Data

SUZANNE L. WEST

The North Carolina Program for Women's Health Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

BRIAN L. STROM

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

CHARLES POOLE

Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

INTRODUCTION

In discussing the quality of data for research, Gordis remarked that epidemiologists have become so enamored with statistical analysis of the data that they have paid too little attention to the validity of the raw data being analyzed with these sophisticated techniques.¹ Although this statement referred to questionnaire data, it applies equally to data generated by abstracting medical records or data from 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 cognitive theories of

autobiographical memory. Most of the chapter presents a literature review of the studies that have evaluated the validity of drug, diagnosis, and hospitalization data and the factors that influence the accuracy of these data. This information will be presented for the three primary information sources available for pharmacoepidemiology studies: questionnaires, administrative databases, and aggregate data sources. 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 PHARMACOEPIDEMIOLOGY 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. This problem is particularly relevant in the clinical situation, where a patient may be treated concurrently or sequentially by several different physicians. In circumstances such as these, the physician may not have a complete past history recorded in the chart 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. Recall may be worse for illnesses and medication use that occurred many years previously. Thus, a physician's ability to diagnose and/or prescribe successfully may be compromised by the patient's recall abilities. A patient's recall ability also plays a role in the success of drug therapy, since a patient needs to recall the physician's instructions for most efficacious drug use. Brody² found that 55 (53%) of 104 patients interviewed immediately after seeing their physician made one or more errors in recalling their therapeutic regimens.

Of particular concern to the subject of this book is the validity of data on drug exposure and disease occurrence, since the typical focus of pharmacoepidemiology research is the association between a drug exposure and the occurrence of disease. Further, many potential confounders of importance in pharmacoepidemiology research (although certainly not all) are either drugs or diseases. Clinicians recognize that patients very often do not know the names of the drugs they are taking currently, and request that they bring their drugs at every visit. In this context, one certainly would not expect a patient to recall past drug use accurately, at least absent 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, still less recall it. Yet, these data elements are crucial to pharmacoepidemiology studies. 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 PHARMACOEPIDEMOLOGY 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 patient recall, it is important to understand the response process in general and the organization of autobiographical 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 interpretation; (ii) search for and retrieval of information to answer the question or to construct an answer; (iii) judgment for formulating the response; and (iv) editing to determine what, if any, information will be disclosed to the interviewer.³⁻⁵ 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 complex or too vague 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.

Cognitive psychologists report that individuals reconstruct memories either as independent events (episodic recall) or as schema, i.e., generalizations based on patterns of events that have a generic quality. Their research also indicates that individuals retrieve some well learned memories directly, but that other memories are constructed based on information that can be recalled and inferences to fill in the gaps.⁶ With passing time, the recall of events (episodic recall) becomes more difficult and biases due to inferential reconstruction are more likely to occur.⁷ In the survey context, the under-reporting of medical conditions and health visits is more widespread as the interval since the event increases.⁸

Memories for recurring similar events often blend together and people have difficulty accurately recalling the individual events. Thus, while the diagnosis and treatment of an infection might be accurately recalled because it occurs at a distinct point in time and is not temporally related to any other health event, Papanicolaou (pap) smears may be represented as generic memories, or schema, because they are done during regularly occurring physical examinations. The accuracy of information reported using schemas depends on which information is sought, how regularly the patterns occur (i.e., physical examinations), and the recall interval between the event and the query. Schema not only make the recall of the specific details of an individual event, such as the date that the last pap smear occurred, more difficult, but can also obscure deviations or exceptions from what typically occurs.⁹ Thus, a respondent may overlook the fact that they missed their last appointment for a pap smear.

There are four basic approaches to providing time cues to respondents in interviews or questionnaires. One is to ask *how long ago* an event took place. A second is to ask *how old* the respondent was at the time of the event. A third is to place the event of interest (e.g., first use of a particular drug) in a previously established chronology of significant *life events* (marriage, menarche, graduation from high school, birth of a child, etc.). A fourth is simply to ask *when* the event of interest occurred. Any of these approaches may be most suitable, depending on the nature of the event of interest, the context of the interview or questionnaire, the age and mental status of the respondent, and other conditions. The fourth approach is the least structured and most open ended, but it can be used effectively in interviews with proper interviewer training. This approach can be implemented so that the respondent essentially makes the choice among the other three options, perhaps on a question-by-question basis, answering “right after my second child was born” for one question, “in 1979” for another, etc. Then, together with the respondent or later, depending on the information offered, the interviewer converts the response to the desired time format (usually calendar time).

As will be evident in the studies described below, researchers have been employing the findings from cognitive research to pharmacoepidemiology studies. The continued and innovative application of these findings can only further improve the methodology for collecting health related data via questionnaire.

THE INFLUENCE OF MEASUREMENT ERROR ON PHARMACOEPIDEMOLOGY 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 that does not exist in the population or, conversely, may fail to detect a risk factor when one truly exists.

In an epidemiologic study, the measure of association is often based on the number of subjects categorized by the cross-classification of presence or absence of disease and exposure. In a study of the association between drug A and disease B, if 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.¹⁰ 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 (OCs) 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 interviewers 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 noncases (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 noncases (i.e., that exposure classification has lower specificity among cases than among noncases) are undoubtedly too simplistic for general reliance.

A difference in the accuracy of recall between cases and noncases could influence the determination of OC exposure and the resulting measure of association. In case-control studies, differential misclassification of exposure can result from recall bias.¹¹ 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.¹²

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,^{13,14} when there are more than two categories of exposure,¹⁵ and for some forms of disease misclassification.^{16,17} 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.^{18,19}

QUANTITATIVE INDICES OF MEASUREMENT ERROR

Three kinds of comparison may be drawn between two (or more) methods 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 method or source 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 schedule. Reliability is not validity, though the term is sometimes used as such.

When different methods or sources of information are compared (e.g., prescription dispensation records and interview responses), but 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 come into use recently.²⁰

For a binary exposure or outcome measure, such as ever versus 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).

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” 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.

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. The proportion of persons classified as having the exposure or outcome who are correctly classified is the predictive value positive. The proportion of persons classified as lacking the exposure or outcome who are correctly classified is the predictive value negative. 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 dispensation 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 variously 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.

Figure 39.1 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. Similarly, there may be substantial gaps between when a prescription is filled and when it is ingested, or that it is used according to the typical prescribed regimen.

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).²¹ 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 underestimates values of the other variable. The high correlation means that the over- or underestimation is systematic and very consistent. When the two measures 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.²² 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 percent agreement and 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.^{10,21} The κ statistic is the percent agreement corrected for chance.²¹ 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 0.81–1.00.²³ Figure 39.2 illustrates the percent agreement and κ calculations for a reliability assessment between questionnaire data and medical record information.

The intraclass correlation coefficient is used to evaluate the reliability of continuous variables.¹⁰ It reflects both the average differences in mean values and 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 is equal to 1.0. Under certain conditions, the intraclass correlation coefficient is

		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 39.1. Formulas for calculating sensitivity and specificity.

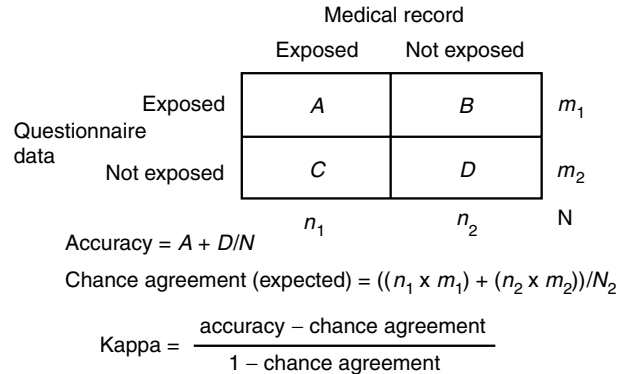


Figure 39.2. Formulas for calculating the percent agreement and κ .

exactly equivalent to Cohen’s weighted κ .²¹ It is difficult 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.* 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.²⁴ 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 intra-class correlation coefficient, percent 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 potential for bias in cohort studies. Likewise, the potential for bias increases in case-control studies, the less prevalent the exposure. 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’s simulations were all done on binary disease and exposure variables. Dosemeci *et al.* 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.¹⁵ 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.¹⁰ For example, if both cases and controls in a case-control study underestimate duration of drug use by an equal percent, 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.²⁴ 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.²⁵ 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.²⁶

For differential misclassification of exposure by disease status (recall bias), Raphael²⁷ 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 statistically control for it. One extremely important way to help make the case for which Raphael has called is to conduct a *sensitivity analysis*.²⁸ Sensitivity analysis is the last line of defense against biases after every effort has been made to eliminate, reduce, or control them in study design, data collection, and data analysis. As used in this context, the meaning of the term "sensitivity" differs from its other epidemiologic meaning as the counterpart to specificity as a measure of classification validity. In a sensitivity analysis, one alters key assumptions or methods reasonably to see how sensitive the results of a study are to those variations. One key assumption, usually implicit, is that the exposure and the outcome in a study have been measured accurately. With estimates from previous research or "guesstimates" from expert experience and judgment, one can modify this assumption and use 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 are highly sensitive to what might seem out of the appropriate context to be minor inaccuracies in exposure or outcome classification.

It has been shown that intuitive judgments, even those of the most highly trained and widely experienced investigators, can be poorly calibrated in such matters. For many years, this kind of assessment has been conducted informally and

qualitatively. 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.

CURRENTLY AVAILABLE SOLUTIONS

OVERVIEW OF APPROACHES USED TO EVALUATE THE ACCURACY OF PHARMACOEPIDEMOLOGY 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.

Etiologic Studies

In the 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 care in recording the information, and on the availability of the medical record. 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. In a recent study, Kirking and colleagues²⁹ found that, in general, prescrip-

tions were poorly documented in the medical record when compared with a pharmacy claims database. Documentation was associated with the number of prescriptions dispensed. Overall, only 39% of prescriptions were documented in the prescriber visit notes of the chart, varying from 34% for patients with 12 or more dispensations in a six-month period to 56% for those with fewer than 12 prescriptions in the same time period.

Monson and Bond used patient pharmacy folders that contained all prescription drug orders as the gold standard for evaluating outpatient medical record completeness.³⁰ Virtually all outpatient medications were obtained from the hospital pharmacy. The medical records for 89% of all patients who received drugs from the pharmacy had documentation that any prescription had been written. Of the 1326 individual prescriptions issued for the 355 patients, 26% were not recorded in the chart, and for only 38% of the 1326 prescriptions did the medical record contain the name, dose, strength, and directions as they appeared on the prescription form. Documentation of therapy was inversely correlated with the number of drugs dispensed, similar results to that of Kirking and colleagues.

Using a different study design that looked at a documentation of a single drug, West *et al.* reported that 89% of the outpatient medical records of patients who were dispensed one fill of one NSAID prescription contained documentation of that prescription.³¹ Age, gender, and whether the NSAID was used for an acute or chronic condition did not influence the likelihood that the drug was documented on the chart. Similarly, Christensen and colleagues reported that 92–95% of antihypertensive medications were noted in the medical record using a database of drug dispensations for comparison.³² Thus, study design appears to influence the findings of medical record documentation, with better results noted for those studies that evaluated only one drug³¹ or one therapeutic class³² compared to those that looked for documentation of all prescriptions in a given period of time.^{29, 30}

The therapeutic class may also affect medical record completeness, with psychotropic medications being poorly documented. Buchsbaum *et al.*

reported that the names of benzodiazepines were more often omitted from the chart than the names of nonbenzodiazepine medications (95% versus 81%, respectively, $p < 0.01$), as were the indications for use (95% versus 57%, respectively, $p < 0.0001$), with 15% of benzodiazepine and only 2% of nonbenzodiazepine prescriptions missing from the chart entirely.³³

Two studies have used computerized files of dispensed prescriptions to evaluate the completeness of drug documentation in the inpatient medical record. Strom *et al.* studied 128 cases of Stevens–Johnson syndrome, looking specifically at only those drugs commonly suspected of causing the syndrome. The patients' inpatient medical records recorded only 50% of the 234 prescriptions for these drugs known to be dispensed to these patients in the 30 days prior to hospitalization, according to computerized Medicaid pharmacy claims files.³⁴ These data indicate that drugs prescribed during an outpatient visit often are not documented in a subsequent inpatient chart, even when the drugs may have caused the disease resulting in the hospitalization. Guess *et al.* reported similarly poor completeness comparing discharge and autopsy reports to the computerized drug file from the Saskatchewan Health Plan.³⁵ In their study of persons with fatal upper gastrointestinal hemorrhage or perforation, NSAID use was mentioned in the discharge or autopsy report for only 31% of cases identified as exposed according to the drug file. In another inpatient medical record study, Lloyd and Rissing reported that the average inpatient medical record contains approximately 2.3 physician errors, such as failure to list treatments or diagnoses that were either treated or affected length of stay.³⁶ They also noted that procedures performed in locations other than the operating room were often omitted from the chart.

In summary, the medical record does not document all medications prescribed for the patient. Record completeness is likely to vary by type of drug, type of chart (outpatient versus inpatient), and the number of drugs prescribed in a given period. This diminishes the usefulness of medical records for verifying self-reported drug exposure.

Methodologic Studies

Exposure confirmation performed as part of etiologic studies are 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, although 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 not recorded on a questionnaire or 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.

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.

In the remainder of the chapter, we will examine the available data on the validity and reliability of data obtained from *ad hoc* questionnaires and automated databases containing patient level data. In each case, we will examine the information available regarding drug exposure data and medical diagnoses separately. We will then end with some conclusions, as well as recommendations for related areas requiring research in the future.

SELF-REPORTED DRUG DATA FROM *AD HOC* QUESTIONNAIRE STUDIES

Accuracy

Concern for the validity of medication data obtained from patients prompted the conduct of several studies, the first of which was published in 1967.³⁷ This study evaluated whether recall of medication use during pregnancy was influenced by the outcome of the birth. Since then, numerous studies have evaluated self-reported medication use for oral contraceptives, for postmenopausal estrogens, and for exposures during pregnancy. Recall accuracy for each of these different types of exposure will be discussed in this section, as well as information from the few studies that report on medications other than hormones or pregnancy related exposures.

Oral Contraceptive Use

Most of the studies concerned with the accuracy of interview data have evaluated how well women remember past use of oral contraceptives (OCs)^{38–44} (see Table 39.1) and hormone replacement therapy^{45–49} (see Table 39.2). Accuracy of self-reported OC and hormone replacement therapy has been assessed as part of etiologic studies by verifying exposure using medical records^{40–44,46–49} or as separate, methodologic studies.^{37–39,45,50} The latter have compared self-reported data to prospectively collected information ascertained as part of cohort studies^{37–39} or from drug dispensation records and files.^{45,50}

First examining the OC studies (Table 39.1), the time between when the exposure occurred and when it was queried (period of recall or recall length) varied from 0 (i.e., current use) to 17 or more years between use and questionnaire date. Two^{38,39} of the five methodologic studies^{37–39,45,50} were of OC recall. The methodologic study by Coulter *et al.*³⁹ employed memory prompts whereas that by Bean *et al.*³⁸ did not. The former study employed two different types of memory aid, one which listed the different OC brands available (memory aid A) and the second which used both

Table 39.1. Studies of self-reported prior oral contraceptive use by period of recall and type of questionnaire administration

Author	Recall period	Questionnaire and sample size	Memory aids	Comparison data source	Findings
Glass <i>et al.</i> , 1974 ⁴¹	3–17 years, median: 5	Personal interview <i>n</i> = 75	No	Medical records	Percent agreement First brand used: 74 Most recent used: 79 Current use: 100
Stolley <i>et al.</i> , 1978 ⁴⁴	2+ years	Personal interview <i>n</i> = 246	List of brand names	Medical records	Percent agreement Most recent OC: 89 κ = 0.93 Previous OC: 63 κ = 0.57 Start date of most recent \pm 1 month: 52 Stop date of most recent \pm 1 month: 74 Total duration \pm 1 month: 36 Total duration \pm 1 year: 77
Bean <i>et al.</i> , 1979 ³⁸	9.1 years	Personal interview <i>n</i> = 160	No	Menstrual and Reproductive History Cohort	Percent agreement First use, exact age: 55 Age \pm 1 year: 88
Adam <i>et al.</i> , 1981 ⁴⁰	Unknown	Self-administered <i>n</i> = 676	No	Medical records	25% exposed by questionnaire 32% exposed by medical charts
Rosenberg <i>et al.</i> , 1983 ⁴³	4–16 years	Personal interview <i>n</i> = 130	Calendar and photos	Medical records	Month-specific agreement as a percent: Duration: 90 Duration and brand: 62 Duration, brand and dose: 54
Coulter <i>et al.</i> , 1986 ³⁹	10–15 years	Personal interview <i>n</i> = 99	A: Brand names listed only B: Calendar and photos	Oxford Family Planning Association Cohort	Percent agreement (All \pm 1 yr) Aid A Aid B Duration 57 69 First use 90 90 Last use 63 79 First brand 69 69 Recent brand 53 63
Nischan <i>et al.</i> , 1993 ⁴²	\leq 20 years	Personal interview <i>n</i> = 758	Calendar and samples	Medical records	Percent agreement Cases Controls First brand used: 77 78 Last brand used: 70 66 First use (\pm 1 year) 72 (r = 0.83) 68 (r = 0.88) Last use (\pm 1 year) 70 (r = 0.89) 69 (r = 0.88) Total duration (\pm 1 month) 24 17 Total duration (\pm 1 year) 65 (r = 0.83) 59 (r = 0.91)

Table 39.2. Studies of self-reported prior estrogen use by period of recall and type of questionnaire administration

Author	Recall period	Questionnaire and sample size	Memory aids	Comparison data source	Findings	κ /correlation
Horwitz et al., 1980 ⁴⁷	Unknown	Telephone interviews <i>n</i> = 324	None	Medical records	Percent agreement Ever use Conventional controls: 80 Alternative controls: 86	κ = 0.58 κ = 0.71
Spengler et al., 1981 ⁴⁹	Unknown	Personal interview <i>n</i> = 153	Samples of estrogens	Medical records	Ever use, pooled cases and controls (percent agreement) – versus private MD's records: 82 – versus hospital records: 80	κ = 0.62 κ = 0.59
Paganini-Hill and Ross 1982 ⁴⁸	Unknown	Personal interview <i>n</i> = 334	None	Medical and pharmacy records	Percent agreement Ever use (versus medical record): 75 Dose (versus medical record): 80 Total duration (by month) ^a	κ = 0.51 κ = 0.61 <i>r</i> = 0.63
Persson et al., 1987 ⁴⁵	≤3 years	Self-administered <i>n</i> = 116	List of brand names	Prescription forms	Percent agreement Month begun, exactly: 85 Name: 85 Dose: 88 Duration:	<i>r</i> = 0.98
Goodman et al., 1990 ⁴⁶	1–11+ years	Personal interviews <i>n</i> = 964	Samples of estrogens	Medical records	Percent agreement Ever use: 87 Duration of use: First use (±3 years): 58	κ = 0.74 <i>r</i> = 0.54 <i>r</i> = 0.57
West et al., 1995 ⁵¹	2–3 years 7–11 years	Telephone interviews <i>n</i> = 103	Pictures of estrogens	Pharmacy database	Recall percent Estrogen name: 78 (95% CI: 70–86) Estrogen name and dose: 26 (95% CI: 17–34) Agreement (%) ±6 months ± 1 year ±2 year First use: 33 33 53 Last use: 29 47 69 Duration: 53 62 68	Note: For those with multiple estrogen exposure, a single estrogen was selected as the target drug for assessing name, dose and dates of use.

^a Results for comparison with physician records only, other records were incomplete.

pictures of OCs and a calendar to record life events (memory aid B).

Overall, the available studies indicate that women accurately remember when they first began using OCs, although the brand names and duration of use are not remembered as well (see Table 39.1). We will present the available data on the accuracy of the date or age of initiation of OC use, duration of use, and brand name in turn.

The two methodologic studies reported that approximately 90% of women were able to recall the age (± 1 year) that they began using OCs, regardless of whether a memory aid was used or the type of memory aid used.^{38,39} Of the studies which verified OC questionnaire data using medical records,^{40–44} the results indicate that women's recall of the exact month and year when they began or stopped OC use was poorly recalled (Table 39.1). The difference in the rates between studies may be attributed to the recall length, which was estimated at up to 10 years for the Stolley *et al.* study and up to 20 years for the Nischan *et al.* study.

The accuracy of self-reported data on the exact months of use varied from a high of 90% in the study by Rosenberg *et al.*, which used picture memory aids⁴³ but measured agreement differently from other authors, to a low of 17% reported by Nischan *et al.*, which provided the respondents with both a calendar and samples of OCs.⁴²

Recall accuracy for OC brand names depended on the order of their use. The first brand used was recalled with reasonable accuracy.^{39,41,42} There were far greater discrepancies among the studies for the accuracy of the most recent OC brand used, but this may be due to how accuracy was defined, by name^{41,44} or by name and dose.^{39,42} It was very difficult for respondents to recall *all* OC brands with their correct dosages. Coulter *et al.* found that only 33% of those provided memory aid A and 48% of those given memory aid B could recall all OCs used for six months or longer.³⁹ Rosenberg *et al.* reported similarly poor results for the recall accuracy of OC duration, brand, and dose.⁴³ In the Nischan *et al.* study,⁴² agreement on duration of brand-specific OC use ranged from 31.4 to 100% ($r = 0.21$ to 0.96). Agreement was best for the most recently

introduced product and lowest for the OCs that had similar brand names.

Postmenopausal Hormone Use

In contrast to the studies of OCs, which evaluated age, duration, and brands used, the studies of postmenopausal estrogen use focused primarily on ever/never use (Table 39.2). There have been two methodologic studies to evaluate the recall accuracy for postmenopausal estrogen use, both of which used pharmacy dispensations for comparison and some type of recall aid. West *et al.* reported a sensitivity of 78% for accurately reporting estrogen name but lower sensitivity for dates and duration of use, with similar results for both recall periods.⁵¹ The specificity for nonsteroidal anti-inflammatory drug or estrogen exposure was 95%. As previously noted, although sensitivity was lower than specificity, the relative importance of these two measures depends on the true prevalence of exposure in the population.

Persson *et al.*⁴⁵ reported similar results to those of West *et al.* for the accuracy of reporting drug name, but much better results for reporting the exact month the estrogen was begun. The difference between the two studies may be due to study design, the questionnaire used to obtain the drug exposure information, or the recency of estrogen use.

The verification studies comparing self-reported estrogen use to medical record documentation reported percent agreements ranging from 75 to 87% and κ ranging from 0.51 to 0.74, indicating moderate to substantial agreement.^{46–49} In particular, the study of Goodman *et al.* verified both use and nonuse of estrogen exposure as part of an etiologic study of breast cancer.⁴⁶ They found that 14% of “users” and 12% of “never users” according to self-report would have been misclassified using the information from medical records for comparison. This interpretation assumes that the medical record is correct and the women's recall is inaccurate. A more likely explanation for the 14% of women whose reported estrogen use could not be confirmed is that their medical record was incomplete, i.e., it failed to document their prescriptions. In fact, this 14% error rate noted by

Goodman *et al.* agrees very well with the results from the studies evaluating the completeness of the medical record for documenting prescriptions.³⁰⁻³²

Use of Nonhormonal Medications

There are several studies assessing the current use of different types of medication, but only a few studies have evaluated how well respondents report past use of medications other than OCs or postmenopausal estrogens (see Table 39.3). The only two methodologic studies^{50,51} that evaluated the recall of non-hormone medications indicated that respondents have great difficulty accurately recalling specific information about drug use. Despite using memory aids and a structured questionnaire, many respondents were unable to recall the name of the nonsteroidal anti-inflammatory drug they had used in the past, and had even more difficulty with the dose, duration, and dates of use.

In the second methodologic study, Van den Brandt *et al.* compared self-report to pharmacy records of dispensed prescriptions, noting that the number of drug dispensations recalled was highest for cardiovascular medications (66%) and poorest for alimentary tract medications (48%).⁵⁰ Recall was influenced by the number of chronically used medications: 71% for one drug, 64% for two drugs, and 59% for three or more drugs, although duration of use was not related to recall. This study could evaluate only recall of current and very recent (within the past two years) medication use. Questionnaire design may have influenced the results of this study. Insufficient space was allocated for the recording of all medications used in the time period under study. If the respondents were unable to record all medications due to space limitations, it would appear that they were unable to recall all medications when this self-reported information was compared to the medications dispensed according to the database.

Four etiologic studies verified the accuracy of self-reported exposure to drugs other than hormones several years in the past, using medical records for comparison.^{40,48,52,53} The study by Adam *et al.*, which compared data collected by self-administered questionnaire to information

from the patient history portion of the medical record, indicated that women greatly under-report antipsychotic agents,⁴⁰ whereas Cotterchio *et al.* reported substantial agreement for antidepressant use, even by drug name.⁵³ Paganini-Hill and Ross showed different agreement rates between self-reported information and medical record documentation depending on the medication.⁴⁸ Of course, discrepancies between the questionnaire data and medical record data can be partially attributed to the incompleteness of the medical records used for comparison.

There are several studies that evaluate the recall accuracy for currently used medications. The first methodologic study conducted by Hulka *et al.* in 1975 looked at how well drugs were reported overall and, for any errors noted, whether patients failed to report drug use or whether the omissions were due to the physicians being unaware of drugs their patients currently use.⁵⁴ The results differed by therapeutic class, but even for the cardiac and diabetic medications, there were many discrepancies between the two sources of drug data. Similar discrepant results were noted for other methodologic studies,⁵⁵⁻⁵⁷ and two etiologic studies^{58,59} assessing the accuracy of self-reported current drug use comparing questionnaire data to medical records, in-home assessments, or pharmacy databases.

Medication Use In Pregnancy

Assessing the frequency and timing of pregnancy associated exposures is important for determining the potential for teratogenic effects. Nine studies addressed how well women recall medication exposures that occurred during the three different trimesters of pregnancy. Recall accuracy for pregnancy related exposures was evaluated by two different methods. Four studies used a pre- and postdelivery questionnaire approach, evaluating the accuracy of reporting exposures that occurred during pregnancy to that recalled at different times postdelivery.^{37,60-62} Two studies used information from the obstetric or hospital records as the gold standard and evaluated self-report using questionnaires,^{63,64} and four others used the interview as the standard and evaluated

Table 39.3. Studies of self-reported drug use other than oral contraceptives and estrogens by period of recall and type of questionnaire administration

Author	Recall period	Questionnaire and sample size	Memory aids	Comparison data source	Drugs	Findings																												
Hulka <i>et al.</i> , 1975 ⁵⁴	Current use	Personal interview <i>n</i> = 357	Drug bottles	Medical records and pharmacy records	Antihistamines Antibiotics Cardiac/congestive heart failure (CHF) CNS Antidiabetics (DM) GI Vitamins	Results were not available by drug. Overall agreement and type of error (in percent) stratified by underlying chronic condition <table border="0"> <tr> <td>Agreement</td> <td>DM</td> <td>CHF</td> </tr> <tr> <td>Patient omits drug(s)</td> <td>40</td> <td>27</td> </tr> <tr> <td>MD not know of drug(s)</td> <td>21</td> <td>22</td> </tr> <tr> <td>Patient or MD omissions</td> <td>18</td> <td>22</td> </tr> <tr> <td></td> <td>21</td> <td>29</td> </tr> </table>	Agreement	DM	CHF	Patient omits drug(s)	40	27	MD not know of drug(s)	21	22	Patient or MD omissions	18	22		21	29													
Agreement	DM	CHF																																
Patient omits drug(s)	40	27																																
MD not know of drug(s)	21	22																																
Patient or MD omissions	18	22																																
	21	29																																
Adam <i>et al.</i> , 1981 ⁴⁰	Unknown	Self-administered <i>n</i> = 676	None	Medical records	Phenothiazine	Reporting errors were less than average for cardiac and antidiabetic drugs whereas CNS, antibiotics, GI had more than the average error rate. <table border="0"> <tr> <td>1% exposed by questionnaire</td> <td></td> </tr> <tr> <td>20% exposed by medical record</td> <td></td> </tr> </table>	1% exposed by questionnaire		20% exposed by medical record																									
1% exposed by questionnaire																																		
20% exposed by medical record																																		
Paganini-Hill and Ross, 1982 ⁴⁸	Unknown	Personal interview <i>n</i> = 334	None	Medical and pharmacy records	Ever use Reserpine Other antihypertensives Barbiturates Thyroid medications	Percent agreement ^a <table border="0"> <tr> <td>86, κ = 0.44</td> </tr> <tr> <td>87, κ = 0.60</td> </tr> <tr> <td>69, κ = 0.38</td> </tr> <tr> <td>83, κ = 0.62</td> </tr> </table>	86, κ = 0.44	87, κ = 0.60	69, κ = 0.38	83, κ = 0.62																								
86, κ = 0.44																																		
87, κ = 0.60																																		
69, κ = 0.38																																		
83, κ = 0.62																																		
Landry <i>et al.</i> , 1988 ⁵⁶	Current use	Self-administered and telephone interviews <i>n</i> = 38	None	In-home assessment	High blood pressure Arthritis/pain Tranquilizers Heart pills Seizures Diabetes pills Insulin Antibiotics Ulcer drugs Blood thinners Thyroid pills Hormones	Percent agreement between home visit and two types of questionnaire <table border="0"> <tr> <td>Self-administered</td> <td>Telephone</td> </tr> <tr> <td>(SA)</td> <td></td> </tr> <tr> <td>90</td> <td>71</td> </tr> <tr> <td>75</td> <td>48</td> </tr> <tr> <td>67</td> <td>42</td> </tr> <tr> <td>89</td> <td>50</td> </tr> <tr> <td>100</td> <td>50</td> </tr> <tr> <td>80</td> <td>100</td> </tr> <tr> <td>100</td> <td>75</td> </tr> <tr> <td>50</td> <td>0</td> </tr> <tr> <td>50</td> <td>25</td> </tr> <tr> <td>100</td> <td>100</td> </tr> <tr> <td>100</td> <td>100</td> </tr> <tr> <td>100</td> <td>50</td> </tr> </table>	Self-administered	Telephone	(SA)		90	71	75	48	67	42	89	50	100	50	80	100	100	75	50	0	50	25	100	100	100	100	100	50
Self-administered	Telephone																																	
(SA)																																		
90	71																																	
75	48																																	
67	42																																	
89	50																																	
100	50																																	
80	100																																	
100	75																																	
50	0																																	
50	25																																	
100	100																																	
100	100																																	
100	50																																	

Continued

Table 39.3 Continued

Author	Recall period	Questionnaire and sample size	Memory aids	Comparison data source	Drugs	Findings
Johnson and Vollmer, 1991 ⁵⁷	Current use	Self-administered <i>n</i> = 83	None	Computerized pharmacy records and in-home assessment	Cardiac agents Diuretics Analgesics Psychotropics Cardiovasculars Anti-inflammatories Anti-asthmatics Sedatives/hypnotics Anti-Parkinson's Thyroid agents Hormones	Sensitivity 0.85 0.77 0.33 0.60 0.75 0.67 0.92 0.60 1.00 0.83 1.00 Percent recall: 48 66 54 61 Specificity 1.00 0.93 1.00 0.98 0.98 0.94 0.98 0.97 1.00 1.00 0.99 PPV ^b 1.00 0.89 1.00 0.92 0.94 0.67 0.92 0.75 1.00 1.00 0.86
Van den Brandt <i>et al.</i> , 1991 ⁵⁰	Up to 2 years	Self-administered <i>n</i> = 207	None	Computerized pharmacy records	Alimentary Cardiovasculars CNS Other	
Kehoe <i>et al.</i> , 1994 ⁵⁸	Current use	Personal interview <i>n</i> = 942	None	Medical records	Insulin Oral hypoglycemics Aspirin Oral steroids Cout medications Antihypertensives	Sensitivity 0.84 0.78 0.73 0.66 0.68 0.88 Specificity 0.99 0.98 0.70 0.96 0.98 0.89
Sandvik and Hunskaar, 1995 ⁵²	Unknown	Personal interview <i>n</i> = 82	None	Medical records	Estrogens Anticholinergics Sympathomemetics	κ (95% CI) 0.58 (0.39–0.77) 0.33 (0.07–0.59) 0.32 (–0.07–0.71)
West <i>et al.</i> , 1995 ⁵¹	2–3 years 7–11 years	Telephone interviews <i>n</i> = 319	Pictures of NSAIDs	Pharmacy database	Nonsteroidal anti-inflammatory drugs (NSAIDs)	Recall percent for any NSAID use: 57 (95% CI: 50–64) Single NSAID dispensed in 12 year period: 41 (95% CI: 32–50) Repeated NSAID use: 85 (95% CI: 76–94) For those with repeated NSAID use, a single NSAID was selected as the target drug for assessing name, dose, and dates of use.

			Agreement (%)	±6 months	±1 year	±2 year
Law et al., 1996 ⁵⁹	Recent use	Personal interview Name only n = 123				
		Computerized pharmacy records and medical records				
		AZT	κ	20	28	51
		ddl	0.80	17	24	42
		ddC	0.95	67	71	80
	Acyclovir	0.74				
	Fluconazole	0.76				
	Ketoconazole	0.64				
Smith et al., 1999 ⁵⁵	Current use	Personal interview and medication inventory n = 55 users 55 nonusers	κ (95% CI)			
		None				
		Serum levels				
		Aspirin	0.16 (0.0–0.32)			
	Propranolol	0.43 (0.27–0.59)				
	Hydrochlorothiazide	0.62 (0.53–0.91)				
	Digoxin	0.94 (0.74–1.0)				
Cotterchio et al., 1999 ⁵³	Ever use of anti-depressants (AD)	Self-administered n = 147 cases with AD 119 controls with AD 57 cases with no AD 57 controls with no AD	AD use by class and by individual type			
		List of 11 most common drugs				
		Medical records				
		AD as a class	0.60 (0.47–0.74)			
		Amitriptyline	0.64 (0.40–0.88)			
		Fluoxetine	0.69 (0.45–0.94)			
		Imipramine	0.28 (–0.24–0.79)			
		Desipramine	0.83 (0.54–1.0)			
		Maproline	0.84 (0.54–1.0)			
		Sertraline	0.64 (0.19–1.0)			
		Doxepin	0.79 (0.38–1.0)			
		Paroxetine	1.0			
		Duration of AD use	0.56 (0.32–0.79)			
	Date AD first taken	0.48 (0.23–0.72)				

^a Results for comparison with physician records only; other records were incomplete.

^b Positive predictive value.

the completeness of obstetric, hospital, and/or pharmacy records.^{37,65-67} Note that in the above discussion, Klemetti and Saxen³⁷ used two techniques to assess the recall accuracy of exposures that occurred during pregnancy. It may be for this reason that their paper has been cited as providing evidence both for and against recall bias.

The studies that used the pre- and postdelivery approach evaluated the completeness of exposure reporting using the predelivery questionnaire as the criterion. All four studies found that, when questioned after delivery, women forget at least some medications that occurred during gestation^{37,60-62} and there is a small tendency to over-report exposures that did not occur during the pregnancy (range: 0.4 to 20%). For example, 34.1% of women failed to report an antibiotic or antibacterial exposure, whereas 0.7% over-reported this exposure.⁶⁰ The recall interval for these four studies spanned from weeks,⁶⁰ to months,^{37,61} to eight years.⁶² Feldman *et al.* and Mackenzie and Lippman reported that recall differed by type of medication. Chronically used medications were recalled more often than acute exposures and salient exposures (those that prompted study initiation) were also more accurately recalled (81%) than were common and less disconcerting exposures (33%).⁶¹ Similarly, Mackenzie and Lippman reported fewer deletions for prescription drugs such as antibiotics, compared with over-the-counter medications such as vitamins, analgesics, and cold preparations.

Whereas Feldman *et al.* found that factors such as maternal age, marital and employment status, and pregnancy outcome did not influence the reporting of pregnancy medication exposures, de Jong⁶² reported better recall in mothers with higher educational attainment and poorer pregnancy outcome (low birth weight, gestational age, or Apgar score). The two other papers^{37,60} did not find recall differences based on pregnancy outcome.

Two studies used the obstetric record as the gold standard to assess the accuracy of self-reported medication histories.^{63,64} These two studies had widely differing times between exposure and self-report. In the study of Tilley *et al.*, women were questioned anywhere from 10 to 30 years after

their pregnancy to elicit information on diethylstilbestrol use.⁶³ Of the women who had been exposed according to their records, 37% either did not recall or denied diethylstilbestrol exposure. Women in the Werler *et al.* study were interviewed during their postpartum hospital stay. Compared with the obstetric record, there was both under- and over-reporting of exposures that differed by the type of exposure *and* by pregnancy outcome. For example, compared to women with favorable pregnancy outcomes, those with an adverse pregnancy outcome more accurately reported using birth control after conception and experiencing urinary tract infections. The Werler *et al.* study supports the potential for recall bias when using nonmalformed infants as the control group for studying teratogenic exposures.

Finally, there are four studies that used the mother's self-reported drug exposure as the criterion and verified reported exposures using medical records and/or pharmacy records.^{37,65-67} Klemetti and Saxen found that only 6% of the mother's reports were incorrect when compared with medical records.³⁷ McCredie *et al.* used medical records to confirm Bendectin[®] exposure in a random sample of 30 mothers who reported use during pregnancy; 57% of the exposures were documented in the charts.⁶⁶ Of the remaining unconfirmed reports, 20% of the mothers' physicians indicated that they often prescribe the drug without recording it on the medical record and 27% of the records were unavailable. Bryant *et al.* reported moderate agreement between self-reported information and medical records for use of any prescription drugs during pregnancy ($\kappa = 0.48$), but as expected, very poor agreement for over-the-counter medications ($\kappa = 0.02$) and vitamin supplements ($\kappa = 0.07$).⁶⁷ de Jong *et al.* reported a 50% confirmation of self-reported medication exposure using pharmacy, general practitioner, and hospital records, with the best verification noted for general practitioner and hospital records combined (69%).⁶⁵ 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.

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. Mitchell *et al.* reported that open ended questions such as "Have you ever used any medications?" yielded 13–45% of the affirmative responses for use of three different medications. The addition of indication-specific questions added an incremental 35–58% affirmative responses concerning exposures. Finally, 20–35% reported drug exposure only when asked medication- (name-) specific questions.¹² This study supports the work by Cottler and Robins,⁶⁸ suggesting that questionnaire design influences the completeness of self-reported psychoactive medication use. Asking medication-specific questions in addition to indication-specific questions increased reported drug use by 26–36%. In particular, the medication-specific questions substantially increased reporting for certain subgroups, including 25–44 year olds, males, African–Americans, and those with eight or more years of education.

Studies have also shown that memory aids such as photographs of medications or calendars improve recall. Using two different types of memory aid, Coulter *et al.* (Table 39.1) noted that recall of total duration of use and date of last use were greatly improved when respondents were provided with pictures of OCs and a calendar compared with a list of OC brands.³⁹ Similarly, Beresford and Coker found that only 29% of women were able to recall the name and dose of the estrogen without pictures, increasing to 71% for those who used pictures to enhance recall.⁶⁹

Recall period, the time between when the exposure occurred and when it is reported influences accuracy of recall. Stolley *et al.* reported that agreement on OC starting date between self-report and medical records was 55% if the OC was used within one month of hospitalization and 45% if the OC was used within the most recent two

years.⁴⁴ Goodman *et al.* reported correlations of 0.7 and 0.4 for agreement on duration of estrogen use for recall periods of 0–11 and 11+ years, respectively, and similar differences in agreement for age at first use for these two recall periods ($r = 0.7$ and $r = 0.5$, respectively).⁴⁶ In the only methodologic study to evaluate recall period, West *et al.*⁵¹ reported that the names of drugs stopped 2–3 years prior to interview were recalled more frequently than those stopped 7–11 years prior to interview (odds of recall = 3.0, 95% CI: 1.6–5.7, and 2.4, 95% CI: 0.9–6.7, for nonsteroidal anti-inflammatory drugs and estrogens, respectively). All three studies indicated greater inaccuracies as more time elapsed between occurrence of exposure and its subsequent reporting.

Further analysis of the methodologic study by West *et al.* indicated that individuals were better able to recall the name of the nonsteroidal anti-inflammatory drug in question (target NSAID) as the number of its dispensations increased: for every four dispensations of the target NSAID, the odds of recalling its name increased by 1.7 (95% CI: 1.3–2.2).⁷⁰ Similarly, as the number of different types of NSAID used increased, there was better recall: for every three different NSAIDs used, there was a 3.6-fold (95% CI: 1.3–9.9) increased odds of recalling the target NSAID name. However, this same study did *not* report similar findings for recalling estrogen use, which indicates that recall accuracy for past drug use and its predictors differs by therapeutic class.

The possibility of recall bias was the motivation for many validation studies of past medication use. Whereas many of the studies that address this issue evaluate recall accuracy for mothers of normal and abnormal infants,^{37, 60, 61, 64, 66, 71} others have assessed differences between cases and controls in case–control studies of non-pregnancy related conditions.^{42–44, 46, 48, 49, 53} Overall, the literature does not strongly support a general or uniform indication of recall bias in either circumstance,^{37, 42, 46, 48, 49, 53, 60, 61, 66, 71} although there are some exceptions.^{43, 44, 64} Werler *et al.* reported differences in recall by both exposure and birth outcome (normal and malformed infants)⁶⁴ and Stolley *et al.* found that cases showed better percent agreement between self-report and medical

records than did controls for starting date of the most recently used OC (61% versus 48%, respectively) and for duration of use (47% versus 31%, respectively).⁴⁴ Rosenberg *et al.* also found similar percent agreements for recalling the duration of past OC use between cases and controls: 94% versus 87%, respectively.⁴³

To date, few studies have evaluated whether demographic and behavioral characteristics influence the recall of past medication use. Cotterchio *et al.* evaluated age, household income, and education as predictors of recall accuracy for reporting ever having used an antidepressant with inconsistent results.⁵³ Calculating percent agreement using the medical record to confirm self-reported data, Goodman *et al.* noted small variations in recall accuracy for any past estrogen use by ethnicity (96% for Japanese ancestry, 91% for non-Japanese ancestry) and education, with more educated women having less accurate recall (90%) than those without a college education (96%).⁴⁶ West *et al.* noted a similar finding for education; sensitivity was 26% for women with some college and 47% for those without any college recalling the NSAID name (odds of recall = 0.4, 95% CI: 0.1–1.3).⁷⁰

Recall accuracy was affected by age in two of four studies evaluated. West *et al.* reported that persons aged 50–65 years of age recalled the NSAID name more accurately than those aged 66–80, odds of recall = 1.8 (95% CI: 1.0–3.4).⁵¹ Van den Brandt *et al.* found better recall accuracy for younger age groups as well, with sensitivities of 65%, 60%, and 58% for those aged 55–59, 60–64, and 65–69, respectively.⁵⁰ Goodman *et al.* did not report age differences in recall accuracy for past estrogen use⁴⁶ and Stolley *et al.*⁴⁴ did not find age differences in the agreement between the user and the prescriber for the name of the OC used most recently. Study design may explain the different results noted; the two studies that reported an age effect were methodologic studies evaluating recall accuracy^{50, 51} whereas the two that reported no age effects^{44, 46} were etiologic studies that reported verification of drug use as a measure of exposure misclassification for the association under study.

Few other demographic factors were evaluated consistently across studies. No differences in recall

accuracy were noted by gender.^{50, 51} Stolley *et al.* reported racial and socioeconomic differences in reporting with whites having better percent agreement than nonwhites (92% versus 83.1%, respectively) and private paying users having better agreement than those receiving public health care funds (91.3% versus 77.1%, $\chi^2 = 6.6$, $p < 0.04$).⁴⁴

Behavioral characteristics such as smoking and alcohol use have been investigated as predictors of recall accuracy.^{46, 70} Nonsmokers had better recall of ever/never use of estrogens (89.8%, $\kappa = 0.79$) than did smokers (82.2%, $\kappa = 0.64$) in the study of Goodman *et al.* study, which was not supported in the work of West *et al.*, which found no relationship between recall accuracy for past NSAID or estrogen use and cigarette smoking. Similarly, West *et al.* reported that current alcohol use was unrelated to recall accuracy for NSAIDs or estrogens.

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

It is apparent from this review, as well as an earlier review by Harlow and Linet,⁷² that most of the work on recall accuracy for past medication use has focused on OCs and replacement estrogens. The results of the studies indicate that both OCs and replacement estrogens 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 photos or lists of brand names. Recall of other drugs appears potentially more problematic. It is difficult to evaluate the recall

accuracy for medications other than OCs and replacement estrogens, because there are so few studies in the literature that address this and there are substantial differences among these studies. 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 medication.

As for data collection techniques for self-reported drug data, it is apparent that providing recall enhancements produces better data. Calendars and photos of drugs augment recall to a greater degree than listing only the brand names of the drugs in question. Using both medication-specific and indication-specific questions also provide more complete drug data. In fact, these techniques, namely photos, calendars, and the two different types of drug question, have become the state of the art for collecting self-reported drug data by personal or telephone interview.

More work is needed to understand the influence of patterns of drug use (duration, drug switching, recall period, totality of past drug exposure) and respondent characteristics (demographic and behavioral) on recall accuracy of past drug use. Interestingly, the two studies that evaluated educational attainment as a predictor of recall accuracy found that those with at least some years of college had *less* accurate recall than those with no more than a high school diploma. In evaluating other demographic factors, gender was unrelated to recall accuracy but the respondent's current age was important in two of the four studies assessing its contribution. The discrepancies among the studies may be caused by differences in study design (methodologic or etiologic investigations) or differences in the drugs evaluated.

In conclusion, longstanding and widespread concern about substantial recall bias as a major concern in case-control studies of medication use appears to be a misapprehension. Further, 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. Thus, epidemiologists who plan to use questionnaire data to investigate drug-disease associations will need to consider which factors may influence recall accuracy in the design of their research protocols.

SELF-REPORTED DIAGNOSIS AND HOSPITALIZATION DATA FROM *AD HOC* STUDIES

Accuracy

Much of the methodologic research on the ability to remember past medical conditions and the factors that influenced this recall has been sponsored by the US National Center for Health Statistics, to determine how accurately chronic conditions were being reported in the US National Health Survey. These were large studies comparing interview data with administrative forms used by the health maintenance organizations participating in the study.^{73,74} Other investigations have verified self-reported medical conditions using medical records as part of case-control or cohort studies.⁷⁵⁻⁸¹ The Baltimore Study⁸² and the study by Heliovaara *et al.*⁸³ differed from the other studies. Persons who had responded to questions regarding chronic illness were then clinically evaluated for presence or absence of the conditions under study.

Just as recall accuracy of past medication use varies by the type of drug, the ability to remember disease conditions varies by disease (Table 39.4). The best reporting was shown for conditions that are specific and familiar, such as diabetes mellitus,^{58,78,79,82-86} hypertension,^{58,75,79,83,84} asthma,^{78,82,83} and cancers such as breast, lung, large bowel, and prostate.^{75,84,87,88} 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-report to clinical evaluation, depending upon the type of condition, there is both under- and over-reporting.^{82,83} Studies

Table 39.4. Studies of self-reported illnesses

Author	Questionnaire and sample size	Comparison data source	Conditions	Findings
Commission on Chronic Illness 1957 ⁸²	Personal interview where one person reported on all those in the household n = 809	Clinical evaluation	Selected conditions ^a Malignant neoplasms Asthma Thyroid disease Diabetes mellitus Psychoses/neuroses Other mental disorders Heart disease Chronic sinusitis Rheumatoid arthritis Osteoarthritis	Percent agreement on presence of disorder (interview and clinical evaluation) 59 99 10 95 17 11 52 72 100 72
Paganini-Hill and Ross, 1982 ⁴⁸	Personal interview n = 334	Medical and pharmacy records	Ever had: Gallbladder disease Hypertension Diabetes Benign breast disease Hysterectomy Oophorectomy	Percent agreement 95 90 96 87 98 93
Tretli <i>et al.</i> , 1982 ⁷⁶	Self-administered n = 12 694	Medical records	Ever had or recently had: MI Stroke Diabetes MI Stroke	Percent agreement on ever having had disorder between questionnaire and records 81 65 66 74 33
Wilcox and Horney, 1984 ⁸¹	Self-administered n = 362	Menstrual and Reproductive Health Cohort	Spontaneous abortion	Percent agreement: 75 Predictors of recall (percent of abortions recalled): Length of pregnancy when abortion occurred: ≤6 weeks: 54 >13 weeks: 93 Recall period: <10 years: 82 ≥20 years: 73

Colditz <i>et al.</i> , 1986 ⁷⁵	Self-administered Sample size varied by disease being validated n = 121 700	Medical records	New report of disease between follow-ups, Nurses Health Study	Percent agreement			
			Breast	99			
			Large bowel	93			
			Thyroid	93			
			Ovary	87			
			Corpus uteri	74			
			Lung	67			
			MI	68			
			Stroke	66			
			Hypertension	100 (only 60% had medical records available to confirm the diagnosis)			
			Fractures	100			
			Spitz <i>et al.</i> , 1988 ⁸⁴	Self-administered n = 72	Medical records	Past medical conditions	Percent agreement
						Hypertension	82
Duodenal ulcer	78						
Coronary artery disease	100						
Diabetes mellitus	80						
Gout	60						
Rheumatoid arthritis	0						
Emphysema	100						
Previous cancers diagnosed two or more years ago:							
Colorectal	100						
Genital tract	93						
Breast	100						
Melanoma	83						
Hematopoietic	100						
Lung	60						
Liver	0						
Bone	0						
Skin	0						
Operations:							
Appendectomy	95						
Hysterectomy	96						
Cholecystectomy	88						

Continued

Table 39.4 Continued

Author	Questionnaire and sample size	Comparison data source	Conditions	Findings
Bush <i>et al.</i> , 1989 ⁷⁹	Self-administered questionnaire <i>n</i> = 107	Medical records	Ever had any of the following: Angina Cancer Cataracts Diabetes Fracture Hypertension MI Stroke	Percent agreement 85 89 76 98 91 86 94 98 κ 0.57 0.72 0.53 0.93 0.71 0.70 0.85
Linnet <i>et al.</i> , 1989 ⁷⁸	Personal interview with some surrogate interviews <i>n</i> = 338	Medical records	Ever had any of the following selected conditions ^a Tuberculosis Chronic sinusitis Diverticulitis Hepatitis Rheumatoid arthritis Asthma	Self Percent agreement 95 80 96 97 95 98 κ 0.49 0.24 0.53 0.59 0.00 0.39 Surrogate Percent agreement 95 87 93 97 93 97 κ 0.38 0.09 0.35 0.27 0.11 0.68
Linton <i>et al.</i> , 1991 ⁹²	Self-administered <i>n</i> = 3588	Ocular examination	Ever been told of having the following: Cataract Macular degeneration	Self-reported disease at the ocular exam compared to results of exam Percent agreement 84 92 Sensitivity 0.46 0.18 Specificity 0.96 0.99
Midthjell <i>et al.</i> , 1992 ⁸⁵	Self-administered <i>n</i> = 507	Medical records	Presence or absence of diabetes	96.4% of diabetes cases and 99.7% of noncases were verified. Approximately 50% of diabetics recalled the year their diabetes was diagnosed.
Nevitt <i>et al.</i> , 1992 ⁸⁰	Self-administered <i>n</i> = 9704	Medical records	Prospective study of nonspine fractures	78% of nonspine fractures were confirmed False positive report rate was 12% for all nonspine fractures Of those with no reported fractures by self-report, medical records confirmed that no fractures occurred Overreporting was lowest for shoulder, wrist/forearm, elbow, ankle, and hip Overreporting was greatest for hand, finger, rib, or face

Author	Study Design	Health examination	Presence of selected chronic conditions ^a	Sensitivity	Specificity	PPV	κ			
Heliovaara et al., 1993 ⁸³	Personal interview n = 7217	Health examination	Arterial hypertension	0.73	0.99	0.90	0.78			
			Coronary disease	0.55	0.97	0.55	0.52			
			Cerebral stroke	0.58	1.00	0.65	0.61			
			Any cardiovascular disease	0.80	0.94	0.80	0.74			
			Bronchial asthma	0.72	0.99	0.59	0.64			
			Pulmonary emphysema	0.23	0.99	0.37	0.28			
			Osteoarthritis	0.34	0.97	0.53	0.37			
			Low back disorder	0.56	0.92	0.44	0.43			
			Any mental illness	0.20	1.00	0.93	0.30			
			Diabetes	0.81	0.99	0.77	0.78			
			Paganini-Hill and Chao, 1993 ⁸⁸	Self-administered Sample size varied by disease being validated Followup 1 n = 9734 Followup 2 n = 8884	Medical records	Development of the following diseases during 2 year followup	Percent agreement			
							1983	1985		
						Hip fracture	85	70	questionnaire	questionnaire
						Heart attack	44	36		
Cancer	80	56								
Colon	70	70								
Rectum	13	25								
Lung	87	57								
Breast	90	35								
Uterine	74	44								
Prostate	57	34								
Bladder	78	47								
Lymphoma/leukemia	65	29								
Kehoe et al., 1994 ⁵⁸	Personal interview n = 942	Medical records				Ever diagnosed with:	Sensitivity	Specificity		
			Arthritis	0.75	0.66					
			Coronary heart disease	0.64	0.96					
			Hypertension	0.91	0.88					
			Diabetes	0.84	0.97					
			Other CVD	0.57	0.82					
			Cancer	0.71	0.89					

Continued

Table 39.4 Continued

Author	Questionnaire and sample size	Comparison data source	Conditions	Findings
Rosamond <i>et al.</i> , 1995 ⁹¹	Personal interview <i>n</i> = 1053	Medical records	Previous occurrence of an acute MI or heart attack	60% of previous acute MIs were confirmed; unconfirmed reports were confused with unstable angina, coronary heart failure, and other conditions
Kriegsman <i>et al.</i> , 1996 ⁸⁶	Personal interviews <i>n</i> = 2380	Medical records (physicians were sent a questionnaire to complete)	Presence or absence of: Nonspecific lung disease Cardiac disease Peripheral atherosclerosis Cerebrovascular disease Diabetes mellitus Malignant neoplasms (excluding nonmelanoma skin cancer) Rheumatoid arthritis and/or osteoarthritis	κ : (95% CI) 0.59 (0.53–0.65) 0.69 (0.65–0.73) 0.38 (0.30–0.46) 0.56 (0.48–0.64) 0.85 (0.81–0.89) 0.66 (0.60–0.72) 0.31 (0.27–0.35)
Law <i>et al.</i> , 1996 ⁵⁹	Personal interview <i>n</i> = 123	Medical records	Previous HIV illnesses Hairy leukoplakia Oral candidiasis Herpes zoster Pneumocystis carinii Kaposi's sarcoma Esophageal candidiasis	κ : 0.14 0.39 0.62 0.82 0.78 0.64
Bergmann <i>et al.</i> , 1998 ⁸⁷	Self-administered <i>n</i> = 65 582	Cancer registry	MD ever diagnosed: All cancer sites Colon Rectum Lung Melanoma Breast Uterus	Sensitivity 0.79 0.85 0.16 0.90 0.53 0.91 0.71 PPV 0.75 0.54 0.71 0.72 0.34 0.85 0.79

Walker et al., 1998 ⁹⁰	Self-administered n = 5787 for heart attack n = 5907 for stroke	Medical record	κ	Prostate	0.90	0.80	False positive rate (%)	
				Bladder	0.67	0.72		
				Leukemia	0.61	0.41		
				Lymphoma	0.64	0.69		
				Diagnosis of heart attack or stroke during 12 year followup				False negative rate (%)
Heart attack	0.74	33	6					
Stroke	0.54	25	11					
Zhu et al., 1999 ⁸⁹	Self-administered Prostate cancer cases, n = 181 Controls, n = 297	Medical record	κ	Diagnoses for selected variables by case-control status and age				
				Age 40-64				
				Vasectomy	Controls 0.53	Cases 0.61	Controls 0.52	Cases 0.71
				Benign prostatic hyperplasia	0.22	0.26	0.46	0.12
				Prostatitis	0.35	0.25	0.37	0.25
				Epididymitis/orchitis	0.31	0.33	0.44	0.22
				UTI	0.43	0.21	0.29	0.45
				Inguinal hernia	0.85	0.69	0.75	0.77
				Kidney stones	0.82	0.78	0.61	0.39
							Age 65-69	

^a Refer to the report for an entire listing.

using medical records to assess recall accuracy for common ailments typically found poor agreement, where under-reporting was often the major cause of the disagreement in some studies^{58,78,84,86} but over-reporting occurred as well, especially for conditions where the diagnostic criteria are less explicit.⁸⁹

Depending on the data source used for comparison, cardiovascular conditions may be over or under-reported as well.^{58,75,76,79,83,84,86,88,90,91} Poor agreement in most studies was due to over-reporting of disease^{75,76,91} but some studies by design could only find over-reporting.⁷⁶ In phase one of a cohort study to ascertain cardiovascular disease occurrence by self-report, only 61 (81%) of 75 myocardial infarctions (MIs), 20 (65%) of 31 strokes, and 29 (66%) of 44 reports of diabetes could be verified by medical record review.⁷⁶ Others noted both under- and over-reporting, depending upon the cardiovascular disease being studied.^{79,86,88,89} Heliovaara *et al.*⁸³ and Spitz *et al.*⁸⁴ noted accurate recall of cardiovascular conditions and Kehoe *et al.*⁵⁸ suggested that their results were due to under-reporting of disease. In most instances of recall error, many who had incorrectly reported MIs and stroke had other conditions which they may have mistakenly understood as coronary heart disease, MI, or stroke, based upon communication with their physician during their diagnostic visits.^{76,88,90,91}

Only two studies assessed the recall accuracy for mental illnesses, both comparing interview to clinical evaluation.^{82,83} The results indicated poor agreement between the two data sources with under-reporting as the primary reason for poor agreement. It is unclear from these studies whether the reason for under-reporting is an unwillingness on the part of the respondent to admit to mental illness or whether the conditions were actually under-diagnosed.

There have been only three studies that evaluated reporting of cataracts, two assessing presence of cataract by clinical examination^{82,92} and the third using medical record review for comparison.⁷⁹ Agreement was best in the study by Bush *et al.* whereas the studies that used clinical assessments typically reported poor agreement. Similar

to the evaluation of mental illnesses, the question remains, could the under-reporting be due to under-diagnosis?

Finally, fractures were evaluated in four studies, all of which used medical records for comparison. The overall results from the studies indicated good agreement although the one methodologic study of fracture incidence indicated a slight tendency for over-reporting of hand, finger, rib, or facial fractures.⁸⁰

Although menarche and menopause are not medical conditions *per se*, the age at which they occur is often of interest in pharmacoepidemiology studies. Using data from the Menstrual and Reproductive Health Study which had recall periods ranging from 17 to 53 years (mean 33.9 years), Bean *et al.* found that the exact age of menarche was recalled by 59%, and age within one year was recalled by 90%.³⁸ Similarly for menopause, 45% of women were able to report their exact age at natural menopause and 75.5% reported age within one year. The percent agreements were 55.6% and 83.4%, respectively, for surgical menopause. The recall lengths for menopause were 7.6 years and 10.6 years for natural vs. surgical menopause, respectively. The lower percent agreement for age at which natural menopause occurred compared to that for surgical menopause may be attributed to the gradual occurrence of natural menopause compared to the definitive nature of hysterectomy.⁹³

Influences on Accuracy

The reporting of a medical condition during an interview is influenced by several factors, including the type of condition as well as the patient'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.^{82,94} As a result, conditions considered sensitive are likely to be under-reported 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. Of those with current restrictions on food or beverage due to medical problems, 64.2% reported chronic conditions that were confirmed in medical records, compared with 58.2% for those without these restrictions.⁷³ Similarly, 71.2% of those who had restrictions on work or housework reported chronic conditions, versus 45.9% for those who did not have these restrictions.⁷³ The major determinant of recall for spontaneous abortions was the length of the pregnancy at the time the event occurred.⁸¹ Fifty-four percent recalled a spontaneous abortion occurring within the first six weeks of pregnancy, whereas 93% remembered those occurring more than 13 weeks into the pregnancy.

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.^{73, 74, 81, 95, 96} For peptic ulcer hospitalizations, 90% of those hospitalized 2–6 months prior to interview remembered the date of admission, compared with 80% for those hospitalized 7–18 months prior to interview and 60% for those hospitalized 5 or more years prior to interview.⁹⁶ For reporting of diagnoses, the longer the interval between the date of the last medical visit for the condition and the date of interview, the poorer the recall was for that condition.^{73, 74} Ninety-one percent of the conditions requiring a visit within the week before interview were recalled, compared to 76% for conditions requiring visits 2–4 weeks prior. The numbers were 54 and 41% for conditions requiring visits 6 months prior and one year prior, respectively.⁷³

One of the limitations of the National Center for Health Statistics study⁷³ is that it only assessed recall of conditions occurring in the past year. An unanswered question is the ability of respondents to recall conditions that were diagnosed and resolved more than one year previously. According to Wilcox and Horney,⁸¹ 82% of women were able to recall a spontaneous abortion that had occurred in the past 10 years, but only 73% recalled those that occurred 20 or more years previously. While not large, can these differences in recall be

explained by age, recall interval, a cohort effect, or some intertwining of all three? What may have been considered “sensitive” by one generation may not be considered as such by the subsequent generation. Further, terminology changes over time with prior generations using the nomenclature “miscarriages” whereas more recent generations use “spontaneous abortions.” Despite differences in recall accuracy by recall interval, Wilcox and Horney's results indicate that spontaneous abortions appear to be recalled fairly accurately, probably as a result of their emotional impact.

Perhaps as a result of the emotional stress, lifestyle changes, and potential financial strain, hospitalizations tend to be reported accurately.⁹⁵ There was only a 9% under-reporting of hospitalizations where surgery was performed, compared to 16% of those without a surgical procedure. The under-reporting in those with only a 1 day hospital stay was 28%, compared with 11% for 2–4 day stays and approximately 6% for stays lasting 5 or more days.

There is also consensus that the type of surgery is remembered accurately.^{77, 96, 97} Coulter *et al.* reported that 90% of the surgeries reported during interview were confirmed by general practitioner records. For the remaining 10%, there was a suggestion that the medical record lacked the information.⁹⁷ Recall of the surgery date (± 1 year) was correct for 87.5%. Recall accuracy was very good for hysterectomy and appendectomy,^{48, 78, 84} most likely because these surgeries are both salient and familiar to respondents. Cholecystectomy⁸⁴ and oophorectomy⁴⁸ were not as well recalled and were subject to some over-reporting. The over-reporting noted in the study of Paganini-Hill and Ross may have been due to the potential incompleteness of the medical records used for comparison. For induced abortions, there was only marginal agreement for the occurrence of surgical abortions as noted by records from a managed care organization, with 19% of women under-reporting their abortion history, 35% over-reporting abortions, and 46% reporting accurately according to their medical record.⁹⁸

There has been a thorough evaluation of the influence of demographic characteristics on re-

porting of chronic illnesses, although the results are conflicting. The most consistent finding is that recall accuracy decreases with age,^{58, 86, 87, 92} although this may be confounded by recall interval, or cohort (generational) effects. Whether gender influences recall accuracy is uncertain. Linet *et al.* found men to report better than women, independent of age,⁷⁸ whereas the Baltimore Study found that women reported better than men,⁸² especially in older age groups.⁷³ There was also a suggestion that the gender and age differences depended upon the disease under investigation,⁸² with women over-reporting malignancies and men over-reporting stroke.⁸⁶ Unlike chronic illnesses, no difference was found for reporting of hospitalizations by age or gender.⁹⁵

There was a consistent finding that reporting of both illnesses and hospitalizations was better for whites than for nonwhites,^{73, 78, 82, 95} but the number of nonwhites in each of the studies was relatively small. Udry *et al.* also reported racial differences in reporting accuracy, with nonwhites under-reporting surgical abortions more frequently than whites.⁹⁸

Reporting by educational level was equivocal, with one study showing no difference,⁷⁴ another study indicating better recall for those with less education,⁷³ and four studies suggesting more accurate responses for those with a college education.^{80, 86, 87, 98} With the exception of the study by Nevitt *et al.*,⁸⁰ many studies found that reporting was more complete for self-respondents compared to proxy respondents,^{74, 78, 82, 92} including reporting for hospitalizations where the under-reporting was estimated at 7% for self-respondents and 14% for proxies.⁹⁵ For self-respondents, those with a poor or fair current health status reported conditions more completely than those with good to excellent health status.⁷³

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.⁹⁹ 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. Cannell and colleagues reported that increasing the length of the question improves recall because of the longer question's inherent redundancy and because respondents are provided with more time to develop an answer to the question.¹⁰⁰ However, longer questions could increase the cost of the research and could needlessly tire the respondents.

In summary, whether a person reports an illness during an interview appears to be related to the type of illness, when it occurred, and its saliency, but is less likely to be mediated by demographic characteristics such as age, 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 patient. Although difficult to measure, respondent motivation appears to influence the completeness of reporting as well.^{73, 82, 95}

VALIDITY OF PHARMACOEPIDEMOLOGY DRUG AND DIAGNOSIS DATA FROM COMPUTERIZED DATABASES CONTAINING ADMINISTRATIVE DATA

In addition to conducting *ad hoc* studies to evaluate drug-disease associations, a variety of computerized, administrative databases are available for pharmacoepidemiology research, as reviewed in Chapters 15 to 24. The strengths and limitations of using each of these databases were also discussed in these chapters. 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 service provided and therefore available for analysis (e.g., prescriptions, dental 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 six or more months) (see also Chapter 25). 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.¹⁰¹ Because the 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.¹⁰² For example, completeness of the drug data might vary by income level if persons with higher incomes and drug co-payments choose to obtain their medications at pharmacies not participating in a prescription plan, which is how pharmacy data are collected. Similarly, a bias may be introduced in the association between a drug and a serious adverse drug reaction if some of the 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 dispensations, procedure logs, etc. In a recent review, Rawson and D'Arcy¹⁰³ used the Saskatchewan Health Databases to describe the different validation analyses that could be conducted to evaluate the usefulness of administrative databases for conducting observational studies. These analyses include reviewing the sources noted above, i.e., medical records, billing records, etc., and suggested that pharmacoepidemiologists assess three others 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 compliance and drug ingestion, and that over-the-counter medications are not included. More details about each of the databases are given in Chapters 15 to 24.

Drug Data in Administrative 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 pharmacoepidemiology research (see Chapter 15).

Using the GHC database, West *et al.* found the medical records confirmed that an identical NSAID had been prescribed as was dispensed according to the pharmacy database for 89% of all persons who received only one dispensation of one NSAID in a 12 year period.³¹ Since the chart review for each NSAID was to begin four weeks prior to the dispensation, some of the remaining 11% of NSAIDs not documented in the medical record may have been prescribed more than four

weeks prior to their dispensation. Therefore, the pharmacy database appears to be reasonably reliable if one uses medical records for verification.

Other data also provide information on the completeness of the pharmacy database. According to a survey of GHC patients, more than 90% of prescription medications used for pain management, such as opioids, sedatives/muscle relaxants, and anti-inflammatory drugs, were always filled at GHC pharmacies (see Chapter 15). Preliminary results in a case-control study among 936 postmenopausal women indicate that 96% of the women filled all their prescriptions through GHC pharmacies.¹⁰⁴ Among 762 study subjects treated with antidepressant medications in 1996–1997, only 1.5% reported obtaining antidepressants from a non-GHC pharmacy in the prior 3 months (see Chapter 15).

Copayments were introduced for some GHC plans in 1985 and by 1993, nearly all plans required modest copayments for visits and drugs. Prior to drug copayments, 99% of all prescriptions were filled at GHC pharmacies. In contrast, as of 1986, only 89% of prescriptions for those with copayments were filled at GHC, compared to 93% for those without drug copayments.¹⁰⁵ Data from a subsample of GHC enrollees with and without drug copayments indicated that the copayments did not affect the out-of-plan drug dispensation but did have an effect on drug use, especially for discretionary drugs.¹⁰⁶ Despite the copayments, even in the Medicare population where approximately 50% of enrollees did not have pharmacy benefits, the most recent data indicates that 98% continue to use GHC pharmacies for dispensation of prescription medications (see Chapter 15).

Kaiser Permanente

As discussed in Chapter 16, the Pharmacy Information Management System (PIMS) has been operational since 1994 and contains dispensation information from 108 Kaiser pharmacies. Most of the Health Plan members have a pharmacy benefit (90%) with a \$10 per drug per month co-payment. Because approximately 15–20% of adult members fill at least some of the

prescriptions at non-Kaiser pharmacies, the PIMS is not a complete source of drug dispensing information for all members.

Approximately 90% of Kaiser Permanente Northwest region members have a pharmacy benefit; the remainder who choose to use Kaiser pharmacies are charged at or below current community prices. Computerized drug data are available at Kaiser Permanente Northwest region and most members fill their prescriptions at plan pharmacies (see Chapter 16).

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 (see Chapter 17). Drug data may be missing for the 10% of members without drug benefits, for drugs that cost less than the copayment, or for those who do not submit their drug claim for reimbursement (see discussion on Group Health Cooperative above, for the effects of copayments). 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 18). Approximately 93% of commercial members and most Medicaid members have drug benefits. 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 19, Medicaid databases have been used extensively for pharmacoepidemiology 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. An FDA-funded validation study of one of the Medicaid databases was completed a few years ago, comparing claims data from Michigan and Minnesota to its primary sources, i.e., data from hospitals, physicians, pharmacies, etc.¹⁰⁷ The results of this study indicated that the demographic and drug data appear to be of extremely high quality. Within pre-established limits, year of birth agreed in 94% of sampled patients, and could not be determined from the medical records in another 2.5%; sex agreed in 95% of patients, and could not be determined from the medical records in another 4%; and the date of a pharmacy's dispensing of each drug agreed in 97% of sampled prescriptions.

In addition, because computerized drug utilization review programs are required by law of Medicaid programs (see Chapter 31), thousands of patient-specific alerts each month are sent to the physicians and pharmacists who are providing the patients' medical care. These alerts request that the practitioner verify the accuracy of the billing data, and, if they agree with the basis of the alert, modify their patients' drug therapy regimen to minimize the risks for a possible drug induced illness. Many of the alerts are responded to by the practitioners involved in writing, and few of their responses indicate that the drug data, upon which the alert was generated, were erroneous.

Finally, as noted earlier in this chapter, Strom *et al.* found that, for 128 cases of Stevens–Johnson syndrome, the Medicaid patient's inpatient medical record had only 50% of the 234 prescriptions for drugs suspected of causing the syndrome known to be dispensed according to the computerized Medicaid pharmacy claims files.³⁴ However, these data probably reflect incompleteness of the medical chart, rather than poor quality claims data on drugs.

Of course, as with the other databases, there is no way to evaluate patient compliance with drugs

dispensed, other than examining patterns of refills for chronically used medications, nor are there data on over-the-counter medications.

Saskatchewan Health Plan

The Saskatchewan Health Plan has also been used considerably for pharmacoepidemiology research, as was discussed in Chapter 20. 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 under the program. The information, once verified, is stored permanently on magnetic tape. In addition, on a regular basis, a sample of paid claims is selected and sent to the beneficiaries for confirmation that the service paid for had been provided and that all the information on the claim was correct.

The Drug Prescription 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. In fiscal year 1997–98, approximately 66% of individuals eligible to receive pharmacy benefits actually did so.

Dutch System

In the Netherlands, there is almost universal computerization of pharmacy records enabling the compilation of prescription of drug histories (see Chapter 21). Sickfunds insured patients, i.e., publicly insured individuals, are assigned to a single pharmacy for all reimbursed prescription drugs. The remaining third of the population has coverage by private insurance firms and have the option to use more than one pharmacy. Those who reside in large cities frequently use more than one pharmacy but those who reside in villages typically do not. The investigators using this system believe that the data are of high quality for two reasons.

First, the computerized dispensing records are subject to financial audit, as they are the basis of reimbursement. Second, patients have traditionally used only one pharmacy since there is no economic incentive for between-pharmacy shopping. Comparing currently used prescription drugs as ascertained by home visits to 157 elderly people with those identified by the Dutch pharmacy database, Lau *et al.* reported that 85% of all drugs in the database and 89% of oral drugs were validated, i.e., they were actually used by the study participants.¹⁰⁸

Tayside Medicines Monitoring Unit (MEMO)

As explained in Chapter 22, 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 dispensation 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. The result of this quality control exercise has not been documented.

The UK General Practice Research Database

The information for this database is amassed from general practitioners who have agreed to provide data for research (Chapter 23). The computerized drug file for this data source is based on physician prescribing, not pharmacy dispensations. Thus, a patient 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 compliance are developed. Alternatively, a patient may see a specialist who prescribes a medication. Because this system relies on the prescribing done by general practitioners, specialist prescribed medications would not be available in the database until the patient needs to have it refilled, a responsibility of the patient's

general practitioner. There are two potential drawbacks for using this pharmacy database for pharmacoepidemiology: (i) compliance, 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 patient's general practitioner. However, compliance does not appear to be a major impediment to using the GPRD for research. As noted in Chapter 23, there is 90% concordance between the prescriptions from the GPRD pharmacy database and the UK Prescription Prescribing Authority, indicating that patients do fill most of the prescriptions written by their general practitioners.

Diagnoses and Hospitalizations in Administrative Databases

Unlike the drug data, 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,¹⁰⁹ monitoring that would not typically occur in the outpatient setting.

Systematic errors as a result of diagnostic coding may impact the validity of both inpatient and outpatient diagnostic data. For example, diseases listed in record linked databases are often coded using the International Classification of Disease (ICD) coding system. Poorly defined diseases are difficult to code using the ICD system and "ruleout" codes are unavailable. It is not clear how healthcare plans deal with "ruleout" diagnoses, i.e., are they included in 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. 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 such as Group Health Cooperative or Kaiser Permanente. Lastly, ICD version changes may produce systematic errors, the effects of which were discussed in Chapter 24.

Group Health Cooperative of Puget Sound

The Group Health Cooperative (GHC) inpatient diagnostic database has records for all discharges from GHC owned hospitals as well as those from a GHC operated wing of another community hospital (Chapter 15). 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 have been available as an annual utilization dataset since June 1989.

In a very recent study, Newton *et al.*¹¹⁰ validated the identification of diabetic complications in GHC's automated diagnostic databases with that available from medical charts for 471 randomly selected diabetics. The sensitivity varied by complication, with the highest sensitivities for the most severe conditions: 95.2% for myocardial infarction, 94.4% for amputation, 90.3% for ischemic heart disease, 91.2% for stroke, 79.2% for osteomyelitis, and 73.5% for retinal problems, all of which agreed to within ± 60 days of the database date.

Psaty *et al.* conducted a study to determine the effects of combined estrogen and progestin therapy on the incidence of acute myocardial infarction (MI) in postmenopausal women.¹⁰⁴ They identified cases from two sources at GHC: (i) the computerized discharge abstracts for the GHC hospitals; and (ii) the GHC claims databases, which as described previously, includes the bills for all services provided by non-GHC physicians and healthcare facilities. To determine the validity of the MI diagnoses and the completeness of case ascertainment, two substudies were conducted. For the validation substudy, the clinical history, cardiac enzyme levels, and electrocardiograms for 60 women who presented to GHC hospitals with an acute MI were selected for review. The records

from 40 women who had elevated levels of cardiac enzymes based on the GHC computerized laboratory data and had a cardiac "non-MI" hospitalization were selected for the case ascertainment substudy. These 100 records (60 from substudy 1 and 40 from substudy 2) were reviewed by Psaty and colleagues, blinded to case ascertainment source, in order to classify the events according to accepted criteria for acute MI events. Among the 60 cases from substudy 1, 58 met standard criteria for probable or definite MI. Among the 40 cardiac "non-MI" events from substudy 2, only 3 met criteria for probable or definite MI. The authors concluded that the MI cases in their study were valid and that the case identification methods were reasonably complete at the GHC hospitals.

Kaiser Permanente

Considerable effort has been undertaken to explore the validity of the diagnosis information in Kaiser Permanente in Northern California Kaiser (see Chapter 16). Evaluating the incidence of diarrhea following the use of clindamycin, the investigators identified the frequency of diarrhea diagnoses in medical records compared to that of computer records for all clindamycin users.¹¹² Out of the approximately 300 persons who received clindamycin, there were ten cases of diarrhea recorded in the medical charts but only two cases recorded in the computer records. The difference was because some of the clinics that prescribed clindamycin did not have a diagnosis of diarrhea available on their data recording form. For other patients, it seemed that some physicians were satisfied to record this new symptom in their notes and did not feel the need to list it as a diagnosis.

The investigators also tested the adequacy of the within-Kaiser Permanente followup by submitting a sample of subjects from the pharmacy cohort to the SEER program, searching for cases of cancer that had been missed (see Chapter 16). Altogether, 15.1% of cancers were missing from their files.

Using the Outpatient Summary Clinical Record (OSCR) database, Levin and colleagues identified 11 357 patients with acid-related disorders (ARDs) in a 6 month period from December 1994 to May 1995.¹¹³ They selected 1511 of these patients for

medical record review to confirm the diagnosis. Overall, 83% of diagnoses were verified, with 90% verification for peptic ulcer disease, 88% for gastroesophageal reflux disease, and 71% for gastritis or dyspepsia. Patients whose diagnoses could not be verified by medical record review often had some other type of ARD and only 5.2% with an OSCR diagnosis of ARD had no evidence of ARD according to the medical record.

The OSCR database is a relatively new database for Northern California Kaiser Permanente, having been initiated in 1994. As more studies are conducted using this data resource to identify individuals with specific disorders, we can expect to see additional medical record verification studies to assess its usefulness for pharmacoepidemiology research.

Harvard Pilgrim Health Care

The Harvard Pilgrim Health Care is nearly unique in that epidemiologic analyses use the same automated records that are used by healthcare 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 often 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.^{114,115} For example, in a recent study of asthma, Donahue and colleagues noted a positive predictive value of 86% when comparing the automated and full-text medical record for the availability of an asthma diagnosis.¹¹⁵ Of the patients who were dispensed asthma medications but who did not have an asthma diagnosis according to the automated medical record, 79% did not have an asthma diagnosis according to the full-text records. A similar quality control analysis was conducted for a study of herpes zoster where the age-specific positive predictive value for diagnoses between the automated and the full-text medical records ranged from 86 to 100%¹¹⁴ indicating that the automated

medical records were sufficient for ascertaining the occurrence of herpes zoster in most instances.

UnitedHealth Group

To date, there has only been one study that has formally evaluated the usefulness of UnitedHealth Group data for pharmacoepidemiologic research. Quam *et al.* noted that a combination of medical and pharmacy claims was more productive for identifying hypertensive patients (96%) than medical claims (74%) or pharmacy claims (67%) alone as compared with medical record abstraction.¹¹⁶ 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-9 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 impact the completeness of the diagnosis data available in the databases. For example, in capitated plans, when billing for reimbursement, 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 noted previously, an FDA-funded validation study of one of the Medicaid programs was completed a few years ago, comparing claims data from Michigan and Minnesota to its primary sources, i.e., data from hospitals, physicians,

pharmacies, etc.¹⁰⁷ For medical services, diagnostic agreement to at least three digits of the ICD-9-CM code occurred in only 41%, agreement within a broad diagnostic category in another 16% (i.e., same body system and/or type of illness), no diagnosis was present on the provider record in 12%, a single diagnosis in 3%, and there was no agreement in 28%. Clearly, this study raised important doubts about the validity of the diagnosis data in Medicaid files. However, it is important to recognize that the authors defined agreement simply on the basis of ICD coding. Thus, included as “disagreement” would be situations like a diagnosis of myocardial infarction or chest pain in one of the data sources and a diagnosis of angina pectoris in the other. This is discussed in more detail in Chapter 19.

In the last few years, considerable effort has been expended on exploring the accuracy of the diagnoses in these Medicaid claims files. Some of the results are presented in Chapter 19. Typically, there is 95% agreement for hospital diagnoses between Medicaid claims and discharge diagnoses. The validity of the discharge diagnoses recorded on medical charts is much more uncertain. The validity of laboratory driven diagnoses (e.g., neutropenia) is high. However, for diagnoses that are difficult to make correctly or are defined poorly in the ICD-9-CM 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 19). Of great importance, therefore, because of the increasing societal concerns about confidentiality (see Chapter 26), at the time of this writing (August 1999), most of the Medicaid programs will no longer allow access to medical records.

Saskatchewan Health Plan

Validation substudies were built into the design of several studies using this database (see Chapter 20). Depending upon the condition under study, the results indicated varying levels of confirmation. Thus, validity should be ascertained with each new condition evaluated. There appears to be a very high correlation between the information on

the charts and that coded in the hospital services system. For example, using the computerized records, Guess *et al.* identified 95 hospitalizations with a fatal outcome where gastrointestinal bleeding and/or perforation was coded as a diagnosis. Comparing these hospitalizations to written discharge summaries and autopsy reports, 73 (76.8%) met the study's case definition for these gastrointestinal events.³⁵ Other validation studies of hospital diagnoses using clinical charts have confirmed the very good agreement between these two data sources, supporting the use of the hospital separation database for pharmacoepidemiology (see Chapter 20).

Outpatient data are more problematic to validate, as reimbursement varies with diagnosis, and obtaining medical records is much more difficult and costly (see Chapter 20). A recent study assessed the ability of the database to conduct depression related research by comparing diagnoses from the physician services database to diagnoses abstracted from outpatient charts. The agreement was generally good; there was 77% agreement between the two data sources for depression treated with an antidepressant.¹¹¹

Dutch System

The Dutch system does not have direct access to outpatient diagnosis information at the current time (see Chapter 21). Access to diagnostic data is being obtained in a subset of the database using PHARMO to link drug exposure to hospital data. There is no information available on the validity of such data at the present time.

The Tayside Medicines Monitoring Unit (MEMO)

To date, MEMO contains only hospital diagnostic data, not outpatient diagnoses. Researchers have conducted validation studies of this data, comparing the coded diagnoses with the actual medical chart data (see Chapter 22). Because of the discrepancies noted in these studies, quality control measures have been instituted to improve the diagnostic accuracy of the computerized data.

The 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. Two studies have formally evaluated the usefulness of the GPRD for conducting pharmacoepidemiologic research (see Chapter 23). These studies have focused on an early concern regarding the completeness of the database related to diagnoses made by consultants. Jick and colleagues^{117,118} conducted a validation study to assess whether the consultant diagnoses as documented in letters to the patient's general practitioner actually appeared in the GPRD. The earlier of the two very similar studies indicated that 87% of these diagnoses appeared in the database; the smaller 1992 study showed that 96% of the consultant diagnoses appeared in the database. These two studies left unknown what proportion of the remaining 52% of patients might have seen a consultant who did *not* provide a consultant letter to the patient's general practitioner.

In a similar assessment, Van Staa and Abenheim evaluated whether hospitalizations were reflected in the GPRD.¹¹⁹ Persons could have been selected for study if they had had a hospitalization for hypoglycemia, a hospitalization for conditions other than hypoglycemia, or if they did not have any hospitalizations in the two-year time period. Of the 553 discharge letters for persons with a hypoglycemic diagnosis and the 510 discharge letters for persons with other hospitalizations, 97.5 and 90.0% were noted in the GPRD, respectively. Although persons without a hospitalization in the two-year time period were selected for study, the authors did not state whether any discharge letters were available for this subgroup. Thus, although two validation studies were performed to evaluate the usefulness of the GPRD for pharmacoepidemiology research, there are still unanswered questions that may impact the results of studies performed using this database.

Conclusions

Formal studies are needed to evaluate the validity and completeness of the diagnostic data present in

many of the administrative databases. Until the evaluation research is undertaken, substudies would need to be conducted to ensure that accurate and complete information exists for the association of interest prior to using these databases for research. Publication of the results of such studies with the formal research would establish the credibility of the results and provide assurance that the data are of high quality. Regardless, given the appropriate questions raised about the validity of the diagnoses in these databases, validation of these data through direct access to the relevant medical records is probably needed for virtually all such studies.

THE FUTURE

There is increased awareness in the epidemiologic literature of the need for valid data. Gordis revisited data validity in 1979,¹ an issue that was raised as early as 1958 by Lilienfeld and Graham on the accuracy of self-reported circumcision.¹²⁰ In review, accuracy has typically been assessed by comparing questionnaire data with data from another source that cannot be considered a gold standard. As a result, "verification" rates determined in these studies measure agreement, not validity. Of course, because truly accurate comparison data sources are not readily available (and would be used for research if they were), a measurement of agreement provides the reviewer with some assessment of the validity of the data analyzed in the study, whether it be derived from a questionnaire or from an administrative database.

How do we use what is known about the validity of drug and diagnosis data from questionnaires and computerized administrative databases in planning future research? As this chapter and the data in Table 25.1 of Chapter 25 indicate, there is substantial variability in the sensitivity and specificity of diagnosis and drug data among the different pharmacoepidemiology data resources. Beginning with diagnosis data from *ad hoc* studies using questionnaires, because the data are collected by self-report, these studies may have excellent information on the primary illness of interest, but data for other contributory illnesses

may be more questionable. If the only source of disease information for *ad hoc* studies is the patient completed questionnaire, the literature indicates that the sensitivity and specificity of reported illnesses vary by the type of condition and its embarrassment potential, the terminology used to describe the condition, its impact on lifestyle, and respondent characteristics.

Turning to the validity of drug data from questionnaire studies, it is not clear how well or poorly individuals remember drug exposures other than OCs and replacement estrogens and what influences their recall and reporting of those exposures. Based on the OC and estrogen data, we can expect that significant events will be recalled accurately, such as when a specific therapeutic class was begun and the overall duration of its use. More specifically, the data on OC histories indicate that, although the first OC brand is recalled fairly accurately, changes in brands are not recalled very accurately. To extrapolate this finding to other medications, similar omissions may result in *ad hoc* studies that attempt to obtain complete drug information for therapeutic classes for which considerable drug switching occurs within the class, e.g., antihypertensives or nonsteroidal anti-inflammatory drugs. For example, in the methodologic study by West *et al.*,⁵¹ the sensitivity for recalling any NSAID use was 85% compared to 30% for recalling the name of a specific NSAID. These results suggest that a drug database may be a better source of data for studies that require more precise information on individual drugs and, more specifically, switching within a therapeutic class. In addition, drug databases are very useful for research questions requiring the evaluation of patterns of prescription drug use, either within a class or among several different drug classes.

Those who work with administrative databases are aware of their deficiencies and realize that there needs to be more formal evaluation of the data accuracy. This has not occurred to date, with a few exceptions. Acknowledging the potential for selection bias in the validation study of the Medicaid database (see Chapter 19), Lessler and Harris found that the medication data are complete and accurate, but there are significant

problems with the diagnostic data.¹⁰⁷ The results of this validation effort identified the need to incorporate medical record review into the study protocol for those who used Medicaid data for research purposes.

The findings reported by Lessler and colleagues apply for pharmacoepidemiologic research using most of the administrative databases described in this chapter. Prescription drug exposure information, which is based on dispensation records, is accurate but requires one very important assumption: that if a person fills the prescription, he or she ingests the drug. Because it may be easier to fill the prescription and then decide later whether or not to use it, dispensation does not equate with ingestion. Researchers have dealt with this problem by requiring two or more dispensations of the same drug in a defined time interval to indicate exposure. There is also the possibility that the prescription is filled by one person and used by a friend or relative. Although this does happen, there is no way to identify how often this occurs and for whom.

In 1990, Avorn stated in an editorial in the journal *Epidemiology* that the specification of clinical events on computer processed claims forms is a relatively new aspect of medical care and requirements for documentation are changing rapidly.¹²¹ Third party payers are requiring that physicians justify their services for reimbursement purposes. As a result, Avorn speculated that the validity of claims based data will improve by becoming more detailed. Although the end result will be more accurate descriptions of clinical events, the primary driver is cost control, not research. This may be seen as a positive development by researchers, but there may be negative consequences related to billing according to reimbursement levels; the clinician or administrative staff may code more serious conditions because they result in higher reimbursement.

Since the early 1990s, managed care organizations have been faced with increased pressure to generate clinically informative, patient-level data. The pressure is predominantly from market forces, whereby plans must document quality of care and performance to compete successfully for, and retain, subscribers. Thus, these managed care

organizations are amassing encounter-level data, i.e., visits with associated diagnoses, for external assessments of plan operation and performance, not as an attempt to develop research databases. Plan performance is being measured by "report cards" using indicators such as mammography, cervical screening, cholesterol screening, and immunization rates. The best known of these performance indicators is the Health Plan Employer Data and Information Set (HEDIS). HEDIS requires health plans to capture comprehensive data on all enrolled persons that serve as the denominators for calculating rates; the numerators are derived from the patient-level data. As more managed care plans invest in the development of encounter-level databases that have linkages to enrollment and other services primarily for HEDIS requirements, the end result may be that researchers will have a greater choice of data sources for conducting pharmacoepidemiology studies.

One of the most obvious difficulties identified in the pilot project¹²² was the data inconsistencies between plans. Reimbursement is a prime motivator for clinicians who work in a fee-for-service environment to submit claims data. Most claims systems require documentation of the diagnosis or procedure for each visit although some plans do not require such documentation. Some plans also allow claims processing staff, many of whom have no clinical training, to complete diagnostic codes. Data from capitated or salaried physicians have problems as well. Because the encounter-level data are not linked to payment, there is no requirement for their submission, processing, or more importantly, quality control. The end result is that data from capitated plans are probably incomplete.

The incompleteness of data from capitated plans was documented in a study by Dresser *et al.*¹²³ In describing whether data from automated systems can be used for measuring clinical quality, they noted dramatic differences in the quality measures under review, i.e., cervical, cholesterol, and mammography screening; immunizations; and prenatal care visits as determined by the computerized patient record, manual chart abstraction, and claims based data. In seeking reasons for these discrepancies, the authors suggested that in capitated plans, prenatal care was probably handled as

one event rather than as multiple physician visits, which made it seem as if the women in capitated plans did not receive adequate prenatal care. Similarly, under capitation, pediatric immunizations were administered during "well child visits" and, as a result, were not documented as separate events. The healthcare system is undergoing constant change in an effort to reduce costs and improve access. If these changes are in the direction of capitated systems, the data that we have traditionally used for research may become more variable than it currently is with regard to completeness, threatening its usefulness for conducting pharmacoepidemiology studies.

The future of pharmacoepidemiology rests in thoughtful study design using valid data. However, the accessibility of valid data may be variable as a result of the rapidly changing healthcare system and the demands put upon the system. Researchers must carefully consider which data source will be most appropriate with respect to its sensitivity and specificity for evaluating the study question.¹²⁴ Recently, there has been a greater focus on data validity, as is evident from reviewing the major epidemiology journals, searching for studies that compare two different data sources for comparison of information on key variables. For studies that use computerized claims, it is incumbent on the researcher to justify the usefulness of the database for the research question under study. As epidemiologists, we must remember that the quality of our results is dependent on the validity of our data.

REFERENCES

1. Gordis L. Assuring the quality of questionnaire data in epidemiologic research. *Am J Epidemiol* 1979; **109**: 21-4.
2. Brody DS. An analysis of patient recall of their therapeutic regimens. *J Chron Dis* 1980; **33**: 57-63.
3. Warnecke RB, Sudman S, Johnson TP, O'Rourke D, Davis AM, Jobe JB. Cognitive aspects of recalling and reporting health-related events: Papanicolaou smears, clinical breast examinations, and mammograms. *Am J Epidemiol* 1997; **146**: 982-92.
4. Tourangeau R. Cognitive sciences and survey methods. In Janine TB, Straf ML, Tanur JM,

- Tourangeau R eds, *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Washington, DC: National Academy Press, 1984; 73–100.
5. Willis, GB, Royston P, Bercini, D. The use of verbal report methods in the development and testing of survey questionnaires. *Appl Cognitive Psychol* 1991; **5**: 251–67.
 6. Reder, LM. Strategy selection in question answering. *Cognitive Psychol* 1987; **19**: 90–138.
 7. Brewer WF. Memory for randomly sampled autobiographical events. In Neisser U, Winograd E, eds., *Remembering reconsidered: Ecological and traditional approaches to the study of memory*. Cambridge: Cambridge University Press, 1988; 21–90.
 8. Mathiowetz, NA. Forgetting events in autobiographical memory: Findings from a health care survey. *Proceedings of the Section on Survey Methods Research: American Statistical Association* 1988; 167–72.
 9. Means B, Loftus EF. When personal history repeats itself: Decomposing memories for recurring events. *Appl Cognitive Psychol* 1991; **5**: 297–318.
 10. Kelsey JL, Thompson WD, Evans AS. *Methods in Observational Epidemiology (Monographs in Epidemiology and Biostatistics 10)*. New York: Oxford University Press, 1986.
 11. Sackett DL. Bias in analytic research. *J Chron Dis* 1979; **32**: 51–63.
 12. Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.
 13. Chavance M, Dellatolas G, Lellouch J. Correlated nondifferential misclassifications of disease and exposure: Application to a cross-sectional study of the relation between handedness and immune disorders. *Int J Epidemiol* 1992; **21**: 537–46.
 14. Kristensen P. Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiology* 1992; **3**: 210–15.
 15. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 1990; **132**: 746–8.
 16. Poole C. Exceptions to the rule about nondifferential misclassification [abstract]. *Am J Epidemiol* 1985; **122**: 508.
 17. Rodgers A, MacMahon S. Systematic underestimation of treatment effects as a result of diagnostic testing inaccuracy: implications for the interpretation and design of thromboprophylaxis trials. *Thromb Haemost* 1995; **73**: 167–71.
 18. Thomas DC. Re: “When will nondifferential misclassification of an exposure preserve the direction of a trend?” *Am J Epidemiol* 1995; **142**: 782–3.
 19. Weinberg CR, Umbach DM, Greenland S. Weinberg *et al.* Reply. *Am J Epidemiol* 1995; **142**: 784.
 20. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol* 1993; **137**: 1251–8.
 21. Maclure M, Willet WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987; **126**: 161–9.
 22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–10.
 23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.
 24. Copeland KT, Checkoway H, McMichael AF, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977; **105**: 488–95.
 25. Greenland S, Kleinbaum DG. Correcting for misclassification in two-way tables and matched-pair studies. *Int J Epidemiol* 1983; **12**: 93–7.
 26. Willett W. An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. *Stat Med* 1989; **8**: 1031–40.
 27. Raphael K. Recall bias: A proposal for assessment and control. *Int J Epidemiol* 1987; **16**: 167–70.
 28. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol* 1996; **25**: 1107–16.
 29. Kirking DM, Ammann MA, Harrington CA. Comparison of medical records and prescription claims files in documenting prescription medication therapy. *J Pharmacoepidemiol* 1996; **5**: 3–15.
 30. Monson RA, Bond CA. The accuracy of the medical record as an index of outpatient drug therapy. *J Am Med Assoc* 1978; **240**: 2182–4.
 31. 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.
 32. Christensen DB, Williams B, Goldberg HI, Martin DP, Engelberg R, LoGerfo JP. Comparison of prescription and medical records in reflecting patient antihypertensive drug therapy. *Ann Pharmacother* 1994; **28**: 99–104.
 33. Buchsbaum DG, Boling P, Groh M. Residents’ underdocumentation in elderly patients’ records of prescriptions for benzodiazepine. *J Med Educ* 1987; **62**: 438–40.
 34. Strom BL, Carson JL, Halpern AC, Schinnar R, Snyder ES, Stolley PD *et al.* Using a claims database to investigate drug-induced Stevens–Johnson syndrome. *Stat Med* 1991; **10**: 565–76.
 35. Guess HA, West R, Strand LM, Helston D, Lydick EG, Bergman U *et al.* Fatal upper gastrointestinal hemorrhage or perforation among users and

- nonusers of nonsteroidal anti-inflammatory drugs in Saskatchewan, Canada 1983. *J Clin Epidemiol* 1988; **41**: 35–45.
36. Lloyd SS, Rissing JP. Physician and coding errors in patient records. *J Am Med Assoc* 1985; **254**: 1330–6.
37. Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformations. *Am J Public Health* 1967; **57**: 2071–5.
38. Bean JA, Leeper JD, Wallace RB, Sherman BM, Jagger H. Variations in the reporting of menstrual histories. *Am J Epidemiol* 1979; **109**: 181–5.
39. Coulter A, Vessey M, McPherson K. The ability of women to recall their oral contraceptive histories. *Contraception* 1986; **33**: 127–37.
40. Adam SA, Sheaves JK, Wright NH, Mosser G, Harris RW, Vessey MP. A case-control study of the possible association between oral contraceptives and malignant melanoma. *Br J Cancer* 1981; **44**: 45–50.
41. Glass R, Johnson B, Vessey M. Accuracy of recall of histories of oral contraceptive use. *Br J Prev Soc Med* 1974; **28**: 273–5.
42. Nischan P, Ebeling K, Thomas DB, Hirsch U. Comparison of recalled and validated oral contraceptive histories. *Am J Epidemiol* 1993; **138**: 697–703.
43. 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.
44. 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.
45. 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.
46. Goodman MT, Nomura AM, Wilkens LR, Kolonel LN. Agreement between interview information and physician records on history of menopausal estrogen use. *Am J Epidemiol* 1990; **131**: 815–25.
47. Horwitz RI, Feinstein AR, Stremlau JR. Alternative data sources and discrepant results in case-control studies of estrogens and endometrial cancer. *Am J Epidemiol* 1980; **111**: 389–94.
48. Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 1982; **116**: 114–22.
49. Spengler RF, Clarke EA, Woolever CA, Newman AM, Osborn RW. Exogenous estrogens and endometrial cancer: a case-control study and assessment of potential biases. *Am J Epidemiol* 1981; **114**: 497–506.
50. Van den Brandt PA, Petri H, Dorant E, Goldbohm RA, Van de Crommert S. Comparison of questionnaire information and pharmacy data on drug use. *Pharm Weekbl [Sci]* 1991; **13**: 91–6.
51. West SL, Savitz DA, Koch G, Strom BL, Guess HA, Hartzema A. Recall accuracy for prescription medications: self-report compared to database information. *Am J Epidemiol* 1995; **142**: 1103–12.
52. Sandvik H, Hunskaar S. General practitioners' management of female urinary incontinence. *Scand J Prim Health Care* 1995; **13**: 108–74.
53. Cotterchio M, Kreiger N, Darlington G, Steingart A. Comparison of self-reported and physician-reported antidepressant medication use. *Ann Epidemiol* 1999; **9**: 283–9.
54. Hulka BS, Kupper LL, Cassel JC, Efird R. Medication use and misuse: physician-patient discrepancies. *J Chron Dis* 1975; **28**: 7–21.
55. Smith NL, Psaty BM, Heckbert SR, Tracy RP, Cornell ES. The reliability of medication inventory methods compared to serum levels of cardiovascular drugs in the elderly. *J Clin Epidemiol* 1999; **52**: 143–6.
56. Landry JA, Smyer MA, Tubman JG, Lago DJ, Roberts J, Simonson W. Validation of two methods of data collection of self-reported medicine use among the elderly. *Gerontologist* 1988; **28**: 672–6.
57. Johnson RE, Vollmer WM. Comparing sources of drug data about the elderly. *J Am Geriatr Soc* 1991; **39**: 1079–84.
58. Kehoe R, Wu SY, Leske C, Chylack LT. Comparing self-reported and physician-reported medical history. *Am J Epidemiol* 1994; **139**: 813–8.
59. Law MG, Hurley SF, Carlin JB, Chondros P, Gardiner S, Kaldor JM. A comparison of patient interview data with pharmacy and medical records for patients with acquired immunodeficiency syndrome or human immunodeficiency virus infection. *J Clin Epidemiol* 1996; **49**: 997–1002.
60. Mackenzie SG, Lippman A. An investigation of report bias in a case-control study of pregnancy outcome. *Am J Epidemiol* 1989; **129**: 65–75.
61. Feldman Y, Koren G, Mattice D, Shear H, Pellegrini E, MacLeod SM. Determinants of recall and recall bias in studying drug and chemical exposure in pregnancy. *Teratology* 1989; **40**: 37–45.
62. de Jong PCMP, Berns MPH, van Duynhoven YTHP, Nijdam WS, Eskes TKAB, Zielhuis GA. Recall of medication during pregnancy: validity and accuracy of an adjusted questionnaire. *Pharmacoepidemiol Drug Safety* 1995; **4**: 23–30.
63. Tilley BC, Barnes AB, Bergstralh E, Labarthe D, Noller KL, Colton T *et al.* A comparison of pregnancy history recall and medical records: implications for retrospective studies. *Am J Epidemiol* 1985; **121**: 269–81.

64. Werler MM, Pober BR, Nelson K, Holmes LB. Reporting accuracy among mothers of malformed and nonmalformed infants. *Am J Epidemiol* 1989; **129**: 415–21.
65. de Jong PCM, Prevoo MLL, Zielhuis GA, Roeleveld N, Gabreels F. Accessibility and validity of data on medical drug use during pregnancy collected from various sources. *J Pharmacoepidemiol* 1991; **2**: 45–57.
66. McCredie J, Krickler A, Elliott J, Forrest J. The innocent bystander: doxylamine/dicyclomine/pyridoxine and congenital limb defects. *Med J Aust* 1984; **140**: 525–7.
67. Bryant HE, Visser N, Love EJ. Records, recall loss, and recall bias in pregnancy: a comparison of interview and medical records data of pregnant and postnatal women. *Am J Public Health* 1989; **79**: 78–80.
68. Cottler LB, Robins LN. The effect of questionnaire design on reported prevalence of psychoactive medications. In: Harris LS, ed, *Problems of drug dependence (NIDA Research Monograph)*, DHSS Publication ADM 55-1393. Washington, DC; (1984).
69. Beresford SA, Coker AL. Pictorially assisted recall of past hormone use in case-control studies. *Am J Epidemiol* 1989; **130**: 202–5.
70. West SL, Savitz DA, Koch G, Sheff KL, Strom BL, Guess HA, Hartzema AG. Demographics, health behaviors, and past drug use as predictors of recall accuracy for previous prescription medication use. *J Clin Epidemiol* 1997; **50**: 975–80.
71. Zierler S, Rothman KJ. Congenital heart disease in relation to maternal use of bendectin and other drugs in early pregnancy. *New Engl J Med* 1985; **313**: 347–52.
72. Harlow SD and Linet MS. Agreement between questionnaire data and medical records: The evidence for accuracy of recall. *Am J Epidemiol* 1989; **129**: 233–48.
73. Madow WG. *Interview Data on Chronic Conditions Compared with Information Derived from Medical Records* (Vital and Health Statistics Series 2, No 23). Washington, DC: National Center for Health Statistics, 1967.
74. National Center for Health Statistics. *Health interview responses compared with medical records*. Rockville (MD): National Center for Health Statistics; 1965. Vital and health statistics. Series 2, no. 7.
75. Colditz GA, Martin P, Stampfer MJ, Willet WC, Sampson L, Rosner B *et al.* Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol* 1986; **123**: 894–900.
76. Tretli S, Lund-Larsen PG, Foss OP. Reliability of questionnaire information on cardiovascular disease and diabetes: cardiovascular disease study in Finnmark county. *J Epidemiol Community Health* 1982; **36**: 269–73.
77. Wingo PA, Ory HW, Layde PM, Lee NC, the Cancer and Steroid Hormone Study Group. The evaluation of the data collection process for a multicenter, population-based, case-control design. *Am J Epidemiol* 1988; **128**: 206–17.
78. Linet MS, Harlow SD, McLaughlin JK, McCaffrey LD. A comparison of interview data and medical records for previous medical conditions and surgery. *J Clin Epidemiol* 1989; **42**: 1207–13.
79. Bush TL, Miller SR, Golden AL, Hale WE. Self-report and medical record report agreement of selected medical conditions in the elderly. *Am J Public Health* 1989; **79**: 1554–6.
80. Nevitt MC, Cummings SR, Browner WS *et al.* The accuracy of self-report of fractures in elderly women: evidence from a prospective study. *Am J Epidemiol* 1992; **135**: 490–9.
81. Wilcox AJ, Horney LF. Accuracy of spontaneous abortion recall. *Am J Epidemiol* 1984; **120**: 727–33.
82. Commission on Chronic Illness. *Chronic Illness in the United States. Vol. IV. Chronic Illness in a Large City: the Baltimore Study*. Cambridge: Harvard University Press, 1957; 297–328.
83. Heliövaara M, Aromaa A, Klaukka T, Knekt P, Joukamaa M, Impivaara O. Reliability and validity of interview data on chronic diseases. *J Clin Epidemiol* 1993; **46**: 181–91.
84. Spitz MR, Jueger JJ, Newell GR. The development of a comprehensive, institution-based patient risk evaluation program: II. Validity and reliability of questionnaire data. *Am J Prev Med* 1988; **4**: 188–93.
85. Midthjell K, Holmen J, Bjørndal A, Lund-Larsen PG. Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trøndelag diabetes study. *J Epidemiol Community Health* 1992; **46**: 537–42.
86. Kriegsman DMW, Penninx BWJH, van Eijk JTM, Boeke AJP, Deeg DJH. Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly: A study on the accuracy of patients' self-reports and on determinants of inaccuracy. *J Clin Epidemiol* 1996; **49**: 1407–17.
87. Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol* 1998; **147**: 556–62.
88. Paganini-Hill A, Chao A. Accuracy of recall of hip fracture, heart attack, and cancer: A comparison of postal survey data and medical records. *Am J Epidemiol* 1993; **138**: 101–6.
89. Zhu K, McKnight B, Stergachis A, Daling JR, Levine RS. Comparison of self-report data and

- medical records data: results from a case-control study on prostate cancer. *Int J Epidemiol* 1999; **28**: 409-17.
90. Walker MK, Whincup PH, Shaper AG, Lennon LT, Thomson AG. Validation of patient recall of doctor-diagnosed heart attack and stroke: A postal questionnaire and record review comparison. *Am J Epidemiol* 1998; **148**: 355-61.
 91. Rosamond WD, Sprafka JM, McGovern PG, Nelson M, Luepker RV. Validation of self-reported history of acute myocardial infarction: Experience of the Minnesota Heart Survey Registry. *Epidemiology* 1995; **6**: 67-9.
 92. Linton KLP, Klein BEK, Klein R. The validity of self-reported and surrogate-reported cataract and age-related macular degeneration in the Beaver Dam Eye Study. *Am J Epidemiol* 1991; **134**: 1438-46.
 93. Colditz GA, Stampfer MJ, Willett WC, Stason WB, Rosner B, Hennekens CH *et al.* Reproducibility and validity of self-reported menopausal status in a prospective cohort study. *Am J Epidemiol* 1987; **126**: 319-25.
 94. Jabine TB. *Reporting Chronic Conditions in the National Health Interview Study. Review of Tendencies from Evaluation Studies and Methodological Test (Vital and Health Statistics Series 2, No. 105)*, DHEW Publication PHS 87-1379. Rockville, MD: National Center for Health Statistics, 1987.
 95. National Center for Health Statistics. *Reporting of hospitalization in the Health Interview Survey*. Washington, DC: National Center for Health Statistics; 1965. (Vital and health statistics. Series 2, no. 6).
 96. Corwin RG, Krober M, Roth HP. Patients' accuracy in reporting their past medical history, a study of 90 patients with peptic ulcer. *J Chron Dis* 1971; **23**: 875-9.
 97. Coulter A, McPherson K, Elliott S, Whiting B. Accuracy of recall of surgical histories: a comparison of postal survey data and general practice records. *Community Med* 1985; **7**: 186-9.
 98. Udry JR, Gaughan M, Schwingl PJ, van den Berg BJ. A medical record linkage analysis of abortion underreporting. *Family Planning Perspectives* 1996; **28**: 228-31.
 99. Lessler J, Tourangeau R, Salter W. *Questionnaire design in the cognitive research laboratory*. Hyattsville (MD): National Center for Health Statistics; 1989. (Vital and health statistics. Series 6, no. 1). DHHS Publication No. PHS 89-1076.
 100. Cannell CF, Marquis KH, Laurent A. *A summary of studies of interviewing methodology*. Rockville (MD): National Center for Health Statistics; 1977. (Vital and health statistics. Series 2, no. 69) DHEW Publication No. HRA 77-1343.
 101. Bonito AJ, Farrelly MC, Han J, Lohr KN, Lubalin JS, Lux L. *Assessment of the feasibility of creating a managed care encounter-level database*, RTI Project Report 6703-003. Prepared for Center for Organization and Delivery Studies, Agency for Health Care Policy and Research, Rockville, MD, June 1997, unpublished.
 102. Stergachis AS. Record linkage studies for post-marketing drug surveillance: data quality and validity considerations. *Drug Intell Clin Pharm* 1988; **22**: 157-61.
 103. Rawson NSB, D'Arcy C. *Assessing the validity of diagnostic information in administrative health care utilization data: Experience in Saskatchewan*. *Pharmacoepidemiol Drug Safety* 1998; **7**: 389-98.
 104. Psaty BM, Heckbert SR, Atkins D, Lemaitre R, Koepsell TD, Wahl PW, Siscovick DS, Wagner EH. The risk of myocardial infarction associated with the combined use of estrogens and progestins in post-menopausal women. *Arch Intern Med* 1994; **154**: 1333-9.
 105. Stergachis AS. *Evaluating the quality of linked automated databases for use in pharmacoepidemiology*. In: Hartzema AG, Porta MS, Tilton HH, eds, *Pharmacoepidemiology: an Introduction*, 2nd edn. Cincinnati, OH: Harvey Whitney, 1991; 222-34.
 106. Harris BL, Stergachis A, Ried LD. The effect of drug co-payments on utilization and cost of pharmaceuticals in a health maintenance organization. *Med Care* 1990; **28**: 907-17.
 107. Lessler JT, Harris BSH. *Medicaid Data as a Source for Postmarketing Surveillance Information*, final report. Research Triangle Park, NC: Research Triangle Institute, 1984.
 108. Lau HS, de Boer A, Beuning KS, Porsius A. Validation of pharmacy records in drug exposure assessment. *J Clin Epidemiol* 1997; **50**: 619-25.
 109. Bright RA, Avorn J, Everitt DE. Medicaid data as a resource for epidemiologic studies: strengths and limitations. *J Clin Epidemiol* 1989; **42**: 937-45.
 110. Newton KM, Wagner EH, Ramsey SD, McCulloch D, Evans R, Sandhu N, Davis C. The use of automated data to identify complications and comorbidities of diabetes: A validation study. *J Clin Epidemiol* 1999; **52**: 199-207.
 111. West SL, Richter A, Nennstiel M, Mauskopf JA. *Assessing the Saskatchewan database for outcomes research studies of depression and its treatment*. *J Clin Epidemiol* 2000; **8**: in press.
 112. Friedman GD, Gerard MJ, Ury HK. Clindamycin and diarrhea. *J Am Med Assoc* 1976; **236**: 2498-500.
 113. Levin TR, Schmittiel JA, Kunz K, Henning JM, Henke CJ, Colby CJ, Selby JV. Costs of acid-related disorders to a health maintenance organization. *Am J Med* 1997; **103**: 520-8.

114. Donahue JG, Choo PW, Manson JE, Platt R. The incidence of herpes zoster. *Arch Intern Med* 1995; **155**: 1605–9.
115. Donahue JG, Weiss ST, Goetsch MA, Livingston JM, Greineder DK, Platt R. Assessment of asthma using automated and full-text medical records. *J Asthma* 1997; **34**: 273–81.
116. 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; **131**: 498–507.
117. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resources in the United Kingdom. *Br Med J* 1991; **302**: 766–8.
118. Jick H, Terris BZ, Derby LE, Jick SS. Further validation of information recorded on a general practitioner based computerized data resource in the United Kingdom. *Pharmacoepidemiol Drug Safety* 1992; **1**: 347–9.
119. Van Staa TP, Abenheim L. The quality of information recorded on a UK database of primary care records: A study of hospitalizations due to hypoglycemia and other conditions. *Pharmacoepidemiol Drug Safety* 1994; **3**: 15–21.
120. Lilienfeld AM, Graham S. Validity of determining circumcision status by questionnaire as related to epidemiological studies of cancer of cervix. *J Natl Cancer Inst* 1958; **21**: 713–20.
121. Avorn J. Medicaid-based pharmacoepidemiology: Claims and counterclaims [editorial]. *Epidemiology* 1990; **1**: 98–100.
122. National Committee for Quality Assurance. 1995. Report Card Pilot Project: Key findings and lessons learned: 21 plans' performance profiles. Technical Report, Washington, DC.
123. Dresser MVB, Feingold L, Rosenkranz SL, Coltin KL. Clinical quality measurement: Comparing chart review and automated methodologies. *Med Care* 1997; **35**: 539–52.
124. West, SL. A comparison of data sources for drug exposure ascertainment in pharmacoepidemiologic studies with emphasis on self-reported information. *Pharmacoepidemiol Drug Safety* 1997; **6**: 215–8.

40

Special Methodological Issues in Pharmacoepidemiology Studies of Vaccine Safety

ROBERT T. CHEN

Vaccine Safety and Development Branch, National Immunization Program (E61), Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA

INTRODUCTION

Vaccines are among the most cost-effective of public health interventions.¹ Where vaccination is widely practiced, morbidity and mortality attributable to vaccine preventable diseases have declined considerably.² No vaccine is perfectly safe or effective, however. With high rates of vaccinations and a low incidence of vaccine preventable diseases, adverse events after vaccination are understandably of concern, and have received increasing attention from the medical community and the public.³ Unfortunately, this concern has often affected the stability of vaccination programs.

For example, widespread publicity raising questions about the safety of pertussis vaccine in Sweden, the United Kingdom, and Japan during the 1970s led to fewer pertussis vaccinations, which were followed by epidemics of pertussis.⁴ Similar concerns in the United States during the early

1980s led to more lawsuits, substantial increases in the price of vaccines, the loss of vaccine manufacturers,⁵ and potential deterrence to the development of new vaccines.⁶ As other nations reach high vaccine coverages and lower the rates of diseases, vaccine safety issues may also threaten the stability of their programs. Concerns about potential associations between measles–mumps–rubella (MMR) vaccine and inflammatory bowel disease and autism, and between demyelinating disease and hepatitis B vaccine, have recently affected public acceptance of these vaccines in the United Kingdom⁷ and France,⁸ respectively.

However, as noted in an extensive review in the early 1990s by the Institute of Medicine (IOM) in the United States,^{9,10} current knowledge about vaccine safety is incomplete and the current research capability is limited. Specifically, the IOM identified the following limitations: (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.

The authors of the report concluded that “if research capacity and accomplishments [are] not improved, future reviews of vaccine safety [will be] similarly handicapped.” In subsequent attempts to overcome these gaps and limitation, epidemiology has been vital in providing the scientific methodology for assessing the safety of vaccines.^{3,11} In this chapter, we discuss the major differences between vaccines and other pharmaceutical products in how epidemiology is applied, with respect to both policy and methodology.

CLINICAL PROBLEMS TO BE ADDRESSED USING PHARMACOEPIDEMOLOGY RESEARCH

POLICY ISSUES

Vaccines share many characteristics with other pharmaceutical products, such as their phased development and licensure, but differ fundamentally in many ways. Understanding these differences is important in appreciating the policy context of vaccine safety.

First, a higher standard of safety is generally expected of vaccines. In contrast to most pharmaceutical products, most of which are administered to ill persons for curative purposes, vaccines are generally given to healthy persons to prevent disease. As an extension of the medical maxim “first do no harm,” tolerance of adverse reactions to products given to healthy persons—especially healthy infants—is substantially lower than that to products administered to persons who are already sick. This lower risk tolerance for vaccines may translate into a need to investigate the possible causes of much rarer adverse events following vaccinations than would be acceptable

for other pharmaceutical products. For example, events occurring at $\sim 1/10^5$ – $1/10^6$ doses like acute encephalopathy after whole-cell pertussis vaccine,^{9,12} Guillain–Barré syndrome (GBS) after swine influenza vaccine,¹³ and oral polio vaccine associated paralytic polio (VAPP)¹⁴ are of concern for vaccines, while side-effects are essentially universal for cancer chemotherapy and 10–30% for persons on high dose aspirin therapy experience gastrointestinal symptoms.¹⁵

The cost and the difficulty of studying events increase with their rarity, however. Furthermore, the ability to provide definitive conclusions from epidemiologic studies of rare events decreases. Attributable risks of the order of $1/10^5$ – $1/10^6$ are on the margin of resolution for epidemiologic methods.^{9,16} Perhaps not surprisingly, the bulk of the published literature to date on vaccine safety has been in the form of case reports and case series, rather than controlled studies with adequate power.^{9,10} To assess the possible association between pertussis vaccination and encephalopathy, the British organized a very large case–control study.¹⁷ Enrolled in the study were all children 2 to 35 months of age in England, Scotland, and Wales who were hospitalized for a variety of neurologic illnesses during a 36-month period ($N = 1167$). The finding of a significant association between vaccine and permanent brain damage was based on only seven exposed cases.¹² The validity of this study finding generated much controversy in and out of the courts.^{9,18} Despite considerably more robust data linking GBS with the swine influenza vaccine,¹³ subsequent controversy^{19,20} resulted in a court ordered independent reexamination of the data²¹ and ultimately partial reperformance of the study, confirming the initial findings.²²

A higher standard of safety is also required of vaccines because of the large number of persons who are exposed, frequently compelled to do so by law or regulation for public health reasons.²³ Such requirements were implemented by public health authorities in these countries because vaccine preventable diseases are generally highly infectious (e.g., measles, pertussis). Vaccinations protect individual vaccinees and may also confer protection indirectly to other susceptible persons in the

population, by limiting the spread of disease organisms (so called *herd immunity*). Without such mandates, a “tragedy of the commons” may occur where high vaccine coverage is reached and the individual risk–benefit ratio becomes less than the societal risk–benefit ratio.^{24,25} Persons may attempt to avoid the risks of vaccination while being protected by the herd immunity resulting from others being vaccinated. However, this commons provided by herd immunity may disappear if too many persons avoid vaccination, with the resulting tragedy that outbreaks return.

Due to the need for almost universal exposure to many vaccines, the medical maxim “first do no harm” applies even more in public health than in clinical medicine, where decisions affect many fewer persons. Inadequately inactivated polio vaccine was administered to about 400 000 persons in the “Cutter incident,” resulting in 260 polio cases.²⁶ Other incidents similar in tragedy if not in scope have occurred due to errors in production.²⁷ Recent concerns that polio vaccine contaminated by simian virus 40 may have been received by millions of persons during the 1950s,²⁸ and that some vaccines may have contained gelatin stabilizers produced in cattle infected with bovine spongiform encephalopathy,²⁹ further highlight the importance of ensuring the safety of a relatively universal human directed “exposure” such as immunizations. These concerns are the basis for strict regulatory control of vaccines by the FDA.³⁰

Very high standards of accuracy and timeliness are needed because vaccine safety studies have extremely narrow margins for error. Unlike many classes of drugs for which other effective therapy may be substituted, vaccines generally have few alternative strains or types (oral and inactivated poliovirus vaccines being the best known exception). The decision to withdraw a vaccine or switch between strains may also have wide ramifications. The circumstances surrounding the use and withdrawal of the 1976 “swine influenza” vaccine have been extensively documented,¹⁹ as have the controversy surrounding the safety of whole-cell pertussis vaccines.⁴ In 1992, the United Kingdom withdrew the license of mumps vaccines containing the Urabe strain after studies suggested a high rate

of vaccine associated meningitis.³¹ The manufacturers subsequently withdrew this product worldwide. This left the countries where the Urabe strain had been the sole mumps vaccine licensed without an alternative vaccine.³² Therefore, establishing associations of adverse events with vaccines and promptly defining the attributable risks are critical in placing adverse events in the proper risk–benefit perspective. An erroneous association or attributable risk can undermine confidence in a vaccine and have disastrous consequences for vaccine acceptance and disease incidence. On the other hand, denials of association despite accumulating evidence can backfire.^{33,34}

Because many vaccinations are mandated for public health reasons and because no vaccine is perfectly safe, several countries have established compensation programs for persons who may have been injured by vaccination.³⁵ Accurate assessments of whether adverse events can be caused by specific vaccines is essential to a fair and efficient vaccine injury compensation program.³⁶ In the United States, for example, the Vaccine Injury Table contains the vaccines, adverse events, and intervals after which no-fault decisions are made in favor of the claimants.³⁷ Periodic revisions of the Vaccine Injury Table are necessary to reflect the best scientific information on associations between vaccines and adverse events.

Finally, recommendations for use of vaccines represent a dynamic balancing of risks and benefits. Vaccine safety monitoring is necessary to weigh this balance accurately. 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, as was done with smallpox vaccine³⁸ and with the shift to either inactivated polio³⁹ or sequential inactivated/live oral polio vaccine schedules.¹⁴ Few other vaccine preventable diseases are likely to be eradicated in the near future, however. Most immunizations will therefore be needed indefinitely, with their attendant adverse reactions and potential for loss of public confidence. Research in vaccine safety can help to distinguish true vaccine reactions from coincidental events,^{9,10} estimate their attributable risk,^{12,40} identify risk factors that may permit development of valid contraindications,^{12,41} and, if

the pathophysiologic mechanism becomes known, develop safer vaccines.^{42–46} Equally importantly, such research demonstrates a commitment to reducing disease from all causes, vaccine-preventable and vaccine-induced, and may help to maintain public confidence in immunizations and the credibility of immunization programs.

CLINICAL ISSUES

Vaccines, like other pharmaceutical products, undergo extensive safety and efficacy evaluations in the laboratory, in animals, and in phased human clinical trials before licensure.³⁰ Phase I trials usually number their subjects in the tens and can only detect extremely common adverse events. Phase II trials generally enroll hundreds of subjects. When carefully coordinated, as in the comparative infant DTaP trials, important conclusions such as the relationship between concentration of antigen, number of vaccine components, formulation technique, effect of successive doses, and profile of common reactions (i.e., reactogenicity) can be drawn. Such studies can affect the choice of the candidate vaccine chosen for Phase III.^{47,48}

Sample sizes for phase III vaccine trials are generally larger than those for drugs. In the most extreme example, more than 200 000 vaccinees were enrolled in the famous Francis field trial of inactivated Salk poliovirus vaccine.⁴⁹ More recent conjugate *Haemophilus influenzae* type b vaccine trials have enrolled 30 000–50 000 vaccinees.^{50,51} Nevertheless, sample sizes for phase III vaccine trials are principally based on efficacy considerations. Inferences on safety are drawn to the extent possible based on the sample size ($\sim 10^2$ – 10^5) and the duration of observation (often <30 days).⁴⁷ This usually means that observations of the common local and systemic reactions (e.g., injection site swelling, fever, fussiness) have been possible. Due to the experimental randomized, double-blind, placebo controlled design of clinical trials, inferences on the causal relationship of an adverse event with the vaccine are then relatively straightforward.^{9,10}

Better standardization of safety evaluations in phase III trials is still needed, however, so that safety data across trials and vaccines can be

compared. In the recently completed phase III trials for infant DTaP, a standard case definition ironically was developed for efficacy, but not for safety—the main reason for the development of DTaP.⁵² For example, definitions of high fever across trials varied by the temperature (39.5 versus 40.5 °C), the mode of measurement (oral versus rectal), and time after vaccination measured (48 versus 72 hours).⁵³ Major differences in detected rates of hypotonic–hyporesponsive episodes after the same whole-cell pertussis vaccine used in the Swedish and Italian trials highlight the difficulty of standardizing assessment of rarer events across cultures and health systems, however.⁵⁴ The finding of delayed excess mortality in some recipients of high titer measles vaccine has also raised difficult questions about design of future vaccine trials.^{55,56} Given the need to better appreciate safety of vaccines given universally to healthy babies and the methodologic difficulties of assessing safety postlicensure, some have argued that larger experimental trials may be needed prelicensure to better assess vaccine safety.^{57,58}

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

SIGNAL GENERATION

Because vaccines are biologic rather than chemical in nature, variation in rate of adverse events by manufacturer or even lot might be expected.^{59,60} This variation in biologic response suggests that other unusual adverse events cannot be ruled out. Surveillance systems need to detect such potential new aberrations in a timely manner. Some factors make identification of true signals difficult, however. Many vaccines are administered early in life, at a time when the baseline risk is constantly evolving and may be affected by other perinatal events. Second, if vaccination rates are high, by definition, most persons with adverse medical events will have had a history of vaccination. Distinguishing causal from coincidental events on a case-by-case basis may not be possible.

STANDARD DEFINITIONS AND EVALUATIVE PROTOCOLS

Case definitions can be used at the time of reporting or at the time of analysis to improve specificity. Applying definitions at the time of reporting may reduce the number of reports processed and lower the operating cost (e.g., Canadian Vaccine Associated Adverse Event (VAAE)).⁶¹ The sensitivity of surveillance may be lower and the difficulty of assessing misclassification greater, however. Alternatively, if the reporting form is open ended, this may increase the sensitivity of surveillance but only at the cost of sorting through many nonspecific reports (e.g., US Vaccine Adverse Event Reporting System (VAERS)).^{62,63} Definitions can be applied at the time of analysis. But substantial variation in diagnostic workup and description of events makes classification difficult without additional followup information, which in turn is usually costly.

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., recurrent injection site abscess⁴³ or repeat hair loss with each vaccination)⁶⁴ or a vaccine-specific laboratory finding (e.g., isolation of mumps vaccine virus from the cerebrospinal fluid of a patient with meningitis⁶⁵ or high tetanus titers in persons with severe local reactions)⁶⁶ can be identified. Whenever the adverse event can also occur in the absence of vaccination, epidemiologic studies are necessary to assess whether vaccinated persons are at higher risk than unvaccinated persons. When multiple vaccinations are administered simultaneously, determining whether events are attributable to particular antigens or one of several combinations is frequently difficult if not impossible.

EXPOSURE

Misclassification of exposure status may occur if there is poor documentation of vaccinations. Such

misdocumentation is more likely if there is substantial mobility between healthcare providers. Documentation of exposure status has been fairly good through school age, due to entry requirements linked to vaccinations. Substantial difficulty may be encountered in ascertaining vaccination status in older persons, however. In the United States, recent and likely future increases in number of licensed vaccines, combined with high mobility between immunization providers (up to 25% annually) due to changes in health insurance plans, are leading to a potential confusing maze of vaccination history misclassifications.⁶⁷

For example, even though an infant may have actually received the diphtheria–tetanus–acellular pertussis (DTaP) or the combined diphtheria–tetanus–pertussis–haemophilus influenzae type b (DTPH) vaccine, the immunization card recorder may, due to habit, erroneously record “DTP”. An infant may have started their immunization series with one provider who uses DTaP vaccine primarily, but due to change in parental health insurance, switched to another provider to complete the series, who uses DTPH primarily. Add in the complexity of whether other vaccines such as polio or hepatitis B vaccines are administered simultaneously or not, at different dose series in the schedule and at different ages, and the number of permutations of vaccine exposures that need assessment for potential safety concerns quickly becomes formidable.⁶⁸

OUTCOME

Because the events being assessed are frequently extremely rare (e.g., encephalopathy after pertussis vaccination), identifying enough cases for a meaningful interpretation of study findings can be a major challenge. Even when technically feasible, a study may be logistically infeasible or the findings likely to be too inconclusive to justify the resources. This was the conclusion of an Institute of Medicine committee that evaluated whether the UK’s National Childhood Encephalopathy Study should be replicated in the United States.⁶⁹ The difficulty with adequate study power is further compounded in assessing rare events in populations less frequently exposed (e.g., vaccines given

to travelers or subpopulations with special indications). Studies of Guillain–Barré Syndrome (GBS) after influenza vaccination required the active surveillance of over 20 million persons for several months.^{70,71}

Many adverse events hypothesized to be caused by vaccines are poorly defined clinical syndromes that are diagnoses of exclusion (e.g., encephalopathy,¹² GBS,¹³ chronic fatigue syndrome,⁷² sudden infant death syndrome (SIDS)).⁷³ Our scientific understanding of these diseases is frequently limited in the absence of vaccination, let alone with vaccination. This poor understanding plus the lack of diagnostic tools for these syndromes severely limits clinical and epidemiologic studies of these illnesses. Furthermore, in highly vaccinated populations, risk interval analyses may be the only epidemiologic study design possible (see Solutions—Analyses). Determining the onset of illness is critical in calculating the risk interval. For certain hypothesized vaccine adverse events, there is no known biological mechanism to allow definition of the risk interval. Diseases with insidious or delayed onset such as autism,⁷⁴ inflammatory bowel disease,⁷⁴ and multiple sclerosis⁸ do not permit determination of the risk interval and are therefore also difficult to study.

ANALYSES, CONFOUNDING, AND BIAS

The possibility that vaccines could be responsible for myriad outcomes leads one to consider cohort studies in which events and person times at risk are enumerated in strata formed by various age group and exposure windows. When outcomes are rare, however, cohort studies can be prohibitively expensive, unless all requisite information is automated and linkable.

Because adverse events are rare, studies typically sample the source population of the cases, assess the exposure status of both groups, and use the ratio of exposure odds among the cases and controls to estimate the risk associated with exposure. Because childhood vaccines are generally administered on schedule and children may have developmental dispositions to particular events, the factor age may confound exposure–outcome relations (e.g., diphtheria–tetanus–pertussis (DTP) vaccine and

Table 40.1. Variables determining the extent of bias attributable to confounding in studies of vaccine adverse events

S = risk of AE in unvaccinated children who lack the contraindication ^a
R = true relative risk of AE associated with vaccination
D = relative risk of AE associated with the contraindication
C = proportion of children with the contraindication
V = proportion vaccinated among children without the contraindication
P = proportion vaccinated among children with the contraindication

^aAE = adverse event.

febrile seizures or SIDS).⁷⁵ 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.⁷⁶ Such factors (e.g., low socioeconomic status) may confound studies of vaccine adverse events and lead to underestimates of the true relative risks. The extent of bias introduced by confounding can be examined as a function of six variables (Table 40.1). Relatively little is known about the nature, frequency, and implications of these variables, however. Vaccination rates are generally high in populations in which vaccine adverse events have become a concern. Those who have not been vaccinated may substantially differ in risks for adverse events from the vaccinated population 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.⁷⁶

CURRENTLY AVAILABLE SOLUTIONS

LOGISTICAL APPROACHES

Spontaneous Reporting Systems

Informal or formal passive surveillance or spontaneous reporting systems (SRS) have been the

cornerstone of most vaccine safety monitoring systems, because of their relative low cost of operations.⁷⁷ The national reporting of vaccine adverse events can be done through the same reporting channels as those used for other adverse drug reactions,⁷⁸ as is the practice in France,⁷⁹ New Zealand,⁸⁰ Sweden,⁸¹ and the United Kingdom⁸² (see also Chapters 10 and 11). An increasing number of countries are collecting safety data specific to vaccinations either with a reporting forms and/or surveillance systems different from the drug safety monitoring systems.⁸³ These countries include Australia,⁸⁴ Canada,^{61,85} Denmark,⁸⁶ India,⁸⁷ Italy,⁸⁸ Mexico,⁷⁷ Netherlands,⁸⁹ Sao Paulo State in Brazil,⁹⁰ and the United States.⁶² Vaccine manufacturers also maintain SRS for their products, which are usually forwarded subsequently to appropriate national regulatory authorities.^{30,77}

Because of their importance in infectious disease control, a significant proportion of vaccines in many countries is purchased or administered by national public health authorities. For example, the public sector (federal, state, and local governments) in coordination with the Centers for Disease Control and Prevention (CDCs), purchases over half of the childhood vaccines administered in the US. In many developing countries, the Ministry of Health in conjunction with WHO's Expanded Program on Immunizations (EPI) ad-

ministers almost all vaccines. Potential vaccine adverse events commonly are first reported to the healthcare providers who administered the vaccine. In many countries, such health workers also participate in surveillance for other diseases. These health authorities (e.g., CDCs) therefore commonly lead or collaborate with the vaccine licensure and regulatory agency (e.g., the US FDA) in developing vaccine adverse event reporting systems. A similar model is followed in Canada.⁶¹

The US Experience

The US National Childhood Vaccine Injury Act of 1986 mandated for the first time that health providers report certain adverse events after immunizations (Table 40.2).⁹¹ The Vaccine Adverse Event Reporting System (VAERS) was implemented jointly by the CDC and FDA in 1990 to provide an unified national focus for collection of all reports of clinically significant adverse events, including but not limited to those mandated for reporting.⁶² The creation of VAERS also provided an opportunity to correct some shortcomings of the predecessor CDC Monitoring System for Adverse Events Following Immunizations (MSAEFI) and FDA Adverse Drug Reaction System.⁹²

To increase sensitivity, the VAERS form is designed to permit narrative descriptions of

Table 40.2. Table of reportable events following vaccination, United States

Vaccine/toxoid	Event interval from vaccination
Tetanus in any combination; DTaP, DTP, DTP–HiB, DT, Td, TT	
A. Anaphylaxis or anaphylactic shock	7 days
B. Brachial neuritis	28 days
C. Any sequela (including death) of above events	No limit
D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Pertussis in any combination; DTaP, DTP, DTP–HiB, P	
A. Anaphylaxis or anaphylactic shock	7 days
B. Encephalopathy (or encephalitis)	7 days
C. Any sequela (including death) of above events	No limit
D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert

(continued)

Table 40.2. *Continued*

Vaccine/toxoid	Event interval from vaccination
Measles, mumps and rubella in any combination; MMR, MR, M, R	
A. Anaphylaxis or anaphylactic shock	7 days
B. Encephalopathy (or encephalitis)	15 days
C. Any sequela (including death) of above events	No limit
D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Rubella in any combination; MMR, MR, R	
A. Chronic arthritis	42 days
B. Any sequela (including death) of above events	No limit
C. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Measles in any combination; MMR, MR, M	
A. Thrombocytopenic purpura	30 days
B. Vaccine-strain measles viral infection in an immunodeficient recipient	6 months
C. Any sequela (including death) of above events	No limit
D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Oral polio (OPV)	
A. Paralytic polio	
—in a nonimmunodeficient recipient	30 days
—in an immunodeficient recipient	6 months
—in a vaccine associated community case	No limit
B. Vaccine-strain polio viral infection	
—in a nonimmunodeficient recipient	30 days
—in an immunodeficient recipient	6 months
—in a vaccine associated community case	No limit
C. Any sequela (including death) of above events	No limit
D. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Inactivated polio (IPV)	
A. Anaphylaxis or anaphylactic shock	7 days
B. Any sequela (including death) of above events	No limit
C. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Hepatitis B	
A. Anaphylaxis or anaphylactic shock	7 days
B. Any sequela (including death) of above events	No limit
C. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
<i>Hemophilus influenzae</i> type b	
A. Early-onset Hib disease	7 days
B. Any sequela (including death) of the above events	No limit
C. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert
Varicella	
A. Events described in manufacturer's package insert as contraindications to additional doses of vaccine	See package insert

The Reportable Events Table (RET) reflects what is reportable by law (42 USC 300aa-25) to the Vaccine Adverse Event Reporting System (VAERS) including conditions found in the manufacturers package insert. In addition, individuals are encouraged to report any clinically significant or unexpected events (even if you are not certain the vaccine caused the event) for any vaccine, whether or not it is listed on the RET. Manufacturers are also required by regulation (21CFR 600.80) to report to the VAERS program all adverse events made known to them for any vaccine. Effective 24 March 1997.

adverse events. All persons, including patients or their parents and not just health professionals, are permitted to report to VAERS, especially clinically significant events. (However, as of 2000, <5% of VAERS reports come from parents.) There are no restrictions set on interval between vaccination and onset of illness nor that a patient have medical care to be reported. Annual reminders about VAERS are mailed to physicians likely to administer vaccines. The form is preaddressed and postage paid so that after completion it can be folded and mailed. Report forms, assistance in completing the form, or answers to other questions about the VAERS are available by calling a 24-hour toll-free telephone number (1-800-822-7967) and on the web (www.vaers.org).

A contractor, under CDC and FDA supervision, distributes, collects, codes (using the Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART)),⁹³ and enters VAERS reports into a database. Reporters of selected serious events receive written requests from VAERS (60 days after vaccination and 1 year after vaccination) for information about the patient's recovery. CDC and FDA have on-line access to the VAERS database and focus their efforts on analytical tasks of interest to the respective agencies. These data (minus personal identifiers) are also available to the public. Since its inception in late 1990, approximately 10 000 VAERS reports have been received annually, ~20% of which are defined as serious (death, life-threatening illness, disability, hospitalization).⁹⁴ Due to this volume, followup by a health professional currently occurs on all reports of deaths and only selected serious events of interest.

Other National Experiences

Several other countries also have substantial experience with passive surveillance for vaccine safety. In 1987, Canada developed the Vaccine Associated Adverse Event (VAAE) reporting system.⁶¹ Reporting forms have checkoff boxes for specific events with accompanying case definitions. Provision is also made for an "other" category. To supplement the VAAE, an active, pediatric hospital based surveillance system that

searches all admissions for possible relationships to immunizations known as the Immunization Monitoring Program—Active (IMPACT) has been operational since 1990.⁹⁵ An Advisory Committee on Causality Assessment, consisting of a panel of experts, has also been formed to review the serious VAAE reports.⁹⁶ The Netherlands also convenes an annual panel to categorize their reports, which are then published.⁸⁹ The UK and most members of the former Commonwealth use the "yellow card" system, where a reporting form is attached to officially issued prescription pads.^{78,82} Data on adverse drug (including vaccine) events from about 40 nations are compiled by the WHO Collaborating Center for International Drug Monitoring in Uppsala.⁹⁷ Preliminary efforts are also underway to "harmonize" collection of postlicensure safety data across nations.⁹⁸

A field guide for implementation of monitoring of Adverse Events Following Immunizations (AEFI) has recently been developed by the WHO.⁹⁹ The primary focus is on detection of correctable programmatic errors such as injection site abscesses (suggestive of inadequate sterilization) and development of a rapid response/assessment team for clusters of more serious events (e.g., toxic shock syndrome from contamination of vaccine vials⁸⁷ or deaths from confusing other medications for vaccines).¹⁰⁰ As of 1997, however, only 12 (14%) of 88 national EPIs had such a system in place.^{83,101}

Classifications, Case Definitions, and Evaluative Protocols

Vaccine adverse events can be classified by frequency (common, rare), extent (local, systemic), severity (hospitalization, disability, death), causality, and preventability (intrinsic to vaccine, faulty production, faulty administration). Wilson developed the first classification system with focus on errors of production (e.g., bacterial, viral, toxin contamination) and administration (e.g., nonsterile apparatus).²⁷ A more recent classification divides adverse events after vaccinations into the following groups:^{102,103} (i) *vaccine induced*, due to the intrinsic characteristic of the vaccine preparation and the individual response of the vaccinee,

these events would not have occurred without vaccination (e.g., vaccine associated paralytic poliomyelitis); (ii) *vaccine potentiated*, would have occurred anyway, but were precipitated by the vaccination (e.g., first febrile seizure in a predisposed child); (iii) *programmatic error*, due to technical errors in vaccine preparation, handling, or administration; (iv) *coincidental*, associated temporally with vaccination by chance or due to underlying illness. The distinction between vaccine induced and vaccine potentiated has recently been clarified for DTP and DT vaccine and infantile spasm, perhaps best observed in Figure 40.1.¹⁰⁴

Definitions of certain vaccine adverse events have been developed in Brazil,⁹⁰ Canada,¹⁰⁵ India,⁸⁷ and the Netherlands.⁸⁹ To improve comparability of data across reporting systems, the *Workshop on Standardization of Definitions for Post-Marketing Surveillance of Adverse Vaccine Reactions* was held in October 1991. Definitions for approximately 20 local, central nervous system, and other adverse reactions were adopted by the workshop participants.¹⁰⁵ These case definitions are printed on the Canadian VAAE form as guidance for what should be reported. The proportion of VAAE reports meeting the case definition criteria has increased from 69 to 87%.⁶¹ Alternatively, in a more open reporting system

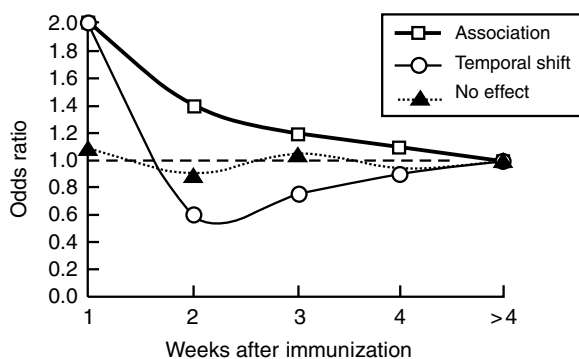


Figure 40.1. Three theoretical models of the temporal relationship between immunization and an adverse effect: (i) *association*, the risk exceeds 1 at all time windows postimmunization; (2) *temporal shift*, the risk exceeds 1 initially then falls below 1 but comes back to 1 eventually, such that the area under the curve above and below 1 is similar; and (3) *no effect*, the risk stays around 1.¹⁰⁴

such as VAERS, these definitions can be applied to reports to develop a case series for further investigation.

Because the simultaneous administration of vaccines is common and multiple adverse events may be reported, the Dutch system further classifies reports as (i) *simple*—a single vaccine injection and a single major reaction, (ii) *compound*—a single vaccine injection and more than one major reaction (each major reaction is counted separately), (iii) *multiple*—>1 vaccine injection in the same person and one major reaction, or (iv) *compound-multiple*—>1 vaccine injection in the same person and >1 major reaction.⁸⁹

To further improve the quality of SRS data and maximize its utility as a potential disease registry, protocols for the clinical evaluation of selected reported serious events of interest (e.g., deaths) can be developed. Such protocols could then be sent to the physicians who report such events to help standardize the evaluation of these patients. In time, a “case series” might then be available for further review that might aid our understanding of the event.⁵²

Assessment of Causality

The formal process of assessing causality of an adverse event and an exposure (e.g., vaccine) is a complex process that can be considered in terms of the answers to three questions: (i) *can it?*, (ii) *did it?*, and (iii) *will it?*.¹⁰⁶ The answer to *can it?* was the focus of the Institute of Medicine reviews.^{9, 10} It is usually based on population level inferences drawn from epidemiologic studies and the following considerations: (i) strength of association, (ii) analytic bias, (iii) biologic gradient/dose-response, (iv) statistical significance, (v) consistency, and (vi) biologic plausibility/coherence.¹⁰⁷

For individual case reports, the *did it?* question is more relevant. If the answer is yes, then *can it?* is also answered in the affirmative. It is natural to suspect vaccine to be the cause when an adverse event occurs in temporal association following vaccination. To base causal inference purely on temporal association, however, is to fall for the logical fallacy of *post hoc ergo propter hoc* (“after this, therefore because of this”).¹⁰ Information

useful for assessing causality in individual case reports include (i) previous general experience with vaccine (e.g., duration of licensure, number of vaccinees, has similar events been observed among other vaccinees or nonvaccinees, do animal models exist to test vaccine as a cause), (ii) alternative etiologies, (iii) individual characteristics of the vaccinee that may increase the risk of the adverse event, (iv) timing of events, (v) characteristics of the event (e.g., laboratory findings), (vi) rechallenge^{108,109} (see also Chapter 32).

When a vaccine *can* cause an adverse event, the *will it?* refers to the probability that an individual will experience the event, or for populations, the proportion that will experience it (i.e., the attributable risk). These data are critical for developing valid contraindications for the individuals at high risk and risk–benefit policy decisions for the population. The *will it?* is usually very difficult to answer, however, as it can only be answered based on epidemiologic studies.¹⁰ Furthermore, the sample sizes of such studies may be large enough to establish whether vaccine can cause a given event but yet inadequate to stratify by subgroups to examine risk factors that can help delineate potential contraindications.

Specific adverse events can usually be said to be caused by a specific vaccine if the event is associated with an unique (i) laboratory finding and/or (ii) clinical syndrome. For example, Urabe mumps vaccine virus was implicated as a cause of aseptic meningitis because mumps virus was isolated from the cerebrospinal fluid (a normally sterile body site) and was shown to be vaccine and not wild strain by genetic sequencing.¹¹⁰ Demonstrations that severe local swelling following tetanus toxoid tended to occur in persons with extremely high levels of circulating antitoxin (due to excessive tetanus boosters) support the proposed mechanism of an Arthus reaction.⁶⁶ Acute flaccid paralysis is almost pathognomonic of polio. In countries such as the US, where wild polio virus is unlikely to be circulating, someone who develops an illness clinically compatible with polio shortly after receipt (or contact with a recipient) of oral polio vaccine is likely to be a case of vaccine associated paralytic polio.^{14,111} Causality can also usually be inferred if a specific clinical finding

occurs after each vaccination (i.e., challenge–rechallenge), as in cases of alopecia after hepatitis B vaccination.⁶⁴

If the adverse event is known to be associated with the wild vaccine preventable disease (e.g., acute arthritis and idiopathic thrombocytopenic purpura after rubella), its association with the attenuated vaccine at a lesser frequency is not surprising.⁹ This relationship is not universal, however, as pregnant women who receive rubella vaccine, unlike those exposed to wild rubella, have not been shown to have illness compatible with congenital rubella syndrome.¹¹² Clustering of events in time after vaccination can also suggest causation if “reporting bias” can be ruled out. Such bias may occur as parents and doctors are more likely to link adverse events with vaccinations the shorter the time interval between the two. Febrile seizures associated with killed bacterial vaccines tend to occur within a day of vaccination while those due to live viral vaccines are delayed by about a week due to viral replication.^{40,113} Onset of GBS after the swine influenza vaccination was delayed up to 6 weeks as autoimmune demyelination is a slower process.⁷¹ The pattern of the risk by time since vaccination may suggest that the relationship to vaccination is more one of temporal shift or triggering of an underlying susceptibility (Figure 40.1).^{104,114}

Unfortunately, most serious reported vaccine adverse events lack these unique features that permit easy inferences on causality. Adverse events such as autism, chronic fatigue syndrome, sudden infant death syndrome, seizures, and GBS either have multiple or as yet unknown etiologies. For these outcomes, vaccination is clearly never the principal “cause” *per se*. Otherwise, given the large number of vaccinations, we would see many more such cases. The question is more whether the association with vaccination can either potentiate the outcome or induce it in a “high risk” subpopulation; or alternatively, is association purely coincidental and vaccination is blamed because it is a highly distinctive, painful, and memorable event usually followed by some true local and systemic vaccine reactions like injection site swelling and fever. For such adverse events, possible link with vaccination is usually based on a

process of elimination, ruling out all other possible causes. Unfortunately, even after this is done, only a relatively unsatisfying nondefinitive conclusion can be drawn on any individual case report because other etiologies may not yet be discovered. The uncertainty in attributing vaccine as a cause of individual cases of illness has led to much confusion, controversy, and litigation.³⁵ With nonunique clinical syndromes or laboratory findings, epidemiologic studies have to be relied upon to ascertain likelihood of association and attributable fraction.

Another approach to causality is to assume that adverse events that occur within particular periods after vaccination are caused by the vaccine, irrespective of whether they were truly causal or just coincidental. This approach to causality is used in some vaccine injury compensation programs to simplify the proceedings.³⁵

In some countries, expert committees of specialists in relevant disciplines (e.g., pediatrics, infectious disease, neurology) review reports. This “global introspection” approach¹¹⁵ has been used in both Canada⁹⁶ and the Netherlands⁸⁹ to classify reports of adverse events in gradations of probable association to vaccination (see also Chapter 32). Classifications are based on the reported symptoms, the interval between vaccination and onset of symptoms, and a set of case definitions. Because opinions of experts play such a major role in this form of causality assessment, the results are less satisfying than results obtained from rigorously conducted scientific studies.

The global introspection method can be improved by the use of branched logic tree algorithms¹¹⁶ or the Bayesian analysis¹¹⁷ (see also Chapter 32). In both, each expert’s degree of belief in the key considerations of the plausibility of vaccine causation is made explicit and measured quantitatively. The algorithm requires the assessor to answer a series of questions which are then scored. The Bayesian analysis calculates the posterior probability of vaccine causation based on applying prior probability that the vaccine can cause the adverse event to the facts of an individual case. Advantages of these approaches include accountability and the possibility of recalculating the probability of causation if the

quality of data improves. Disadvantages include, however, the resources required and the frequent lack of information to construct the prior probabilities. This approach was piloted in a review of MSAEFI cases¹⁰² and used by the Institute of Medicine to review case reports,¹⁰ but has not yet been adopted for routine use.

Signal Detection

Identifying a potential new vaccine safety problem (“signal”) requires a mix of clinical intuition and epidemiologic expertise.¹¹⁸ As indicated above, unusual clinical features and/or clustering by time or space usually tip off that something may be awry. For example, report by a concerned mother of recurrent alopecia after successive hepatitis B vaccinations in her child led to a review of VAERS data that showed several other similar reports.⁶³ No illness other than GBS was reported more commonly in the second and third week than in the first week after swine influenza vaccination, leading to further validation studies.^{13,119} Efforts are under way to explore whether use of artificial intelligence/neural networks may aid signal detection.¹²⁰

Mass immunization campaigns may make it easier to detect adverse events following vaccination than routine immunization activities. The very large number of doses administered over a well defined short time result either in more prominent clusters of vaccine adverse events, or by their absence demonstrate their safety. Surveillance of vaccine adverse events around the time of mass immunization campaigns have therefore been extremely useful in generating signals, either positive (e.g., GBS with swine influenza vaccine,¹³ GBS after oral polio vaccine,¹²¹ allergic reactions after Japanese encephalitis vaccine,¹²² neuropathy after rubella vaccine)¹²³ or negative (e.g., events after meningococcal vaccine,¹²⁴ events after measles–rubella vaccine,¹²⁵ GBS after measles).¹²⁶ Such signals still require validation, however, since some, after more careful scientific studies, turn out to be incorrect (e.g., GBS after oral polio vaccine).¹²⁷ The above situations aside, ascertaining “signals” from spontaneous reports requires taking an epidemiologic approach. Various means

to automate “screening for signals” using spontaneous reports have been tried to date without great success,^{128,129} largely due to inherent methodologic problems of spontaneous reports (see above, and Chapters 10 and 11).

Lessons Learned to Date

Several lessons are beginning to emerge from spontaneous reporting systems such as VAERS.¹³⁰ VAERS has successfully detected unrecognized potential reactions and obtained data to evaluate whether these events are causally linked to vaccines.⁶⁴ VAERS has also successfully served as a source of cases for further investigation of idiopathic thrombocytopenic purpura after MMR,¹³¹ anaphylaxis after MMR,¹³² syncope after immunization,¹³³ and intussusception after rotavirus vaccine.¹³⁴ VAERS has been of great value for answering routine public queries such as “has adverse event *X* ever been reported after vaccine *Y*?” When denominator data on doses are available from other sources, VAERS can be used to evaluate changes in reporting rates over time or when new vaccines replace old vaccines. For example, VAERS showed that after millions of doses had been distributed, reporting rates for serious events such as hospitalization and seizures after DTaP in toddlers was one-third that after DTP.¹³⁵ VAERS is also currently the only surveillance system which covers the entire US population with data available on a relatively

timely basis. It is, therefore, the major means available currently to detect possible new, unusual, or extremely rare adverse events, including whether certain lots of vaccines are associated with unusually high rates of adverse events.^{63,136} VAERS’ predecessor system collected information on personal and family history of convulsions and neurologic illness on all reports. This permitted comparison of the relative proportion of such histories in persons reporting convulsions versus those reporting other nonneurologic adverse events in a crude case–control study.⁴¹

The reporting efficiency or sensitivity of a spontaneous reporting system can be estimated if expected rates of adverse events generated from carefully executed studies are available. A higher proportion of serious events such as seizures that follow vaccinations are more likely to be reported to VAERS than milder events such as rash or delayed events requiring laboratory assessment such as thrombocytopenic purpura after measles–mumps–rubella vaccination (Table 40.3).¹³⁷ Although formal evaluation has been limited, the probability that a serious event reported to VAERS has been accurately diagnosed (i.e., predictive value positive) is likely to be high. Of 26 patients reported to VAERS who developed GBS after influenza vaccination during the 1990–1991 season, whose hospital charts were reviewed by an independent panel of neurologists blinded to immunization status, the diagnosis of GBS was confirmed in 22 (85%).⁷⁰

Table 40.3. Reporting efficiencies for selected outcomes: two passive surveillance systems for vaccine adverse events, US

Adverse event	Vaccine	Reporting efficiency (%)		
		MSAEFI ^a	VAERS ^a (overall)	VAERS ^a (public sector)
Vaccine associated polio	Oral polio vaccine (OPV)	72	68	^b
Seizures	Diphtheria–tetanus–pertussis (DTP)	42	24	36
Seizures	Measles–mumps–rubella (MMR)	23	37	49
Hypotonic–hyporesponsive episodes	DTP	4	3	4
Rash	MMR	<1	<1	<5
Thrombocytopenia	MMR	<1	4	<1

^aMSAEFI = Monitoring System for Adverse Events Following Immunizations; VAERS = Vaccine Adverse Event Reporting System.

^bPublic and private sector information is missing on these cases.

Source: ¹³⁷

Despite the above uses, spontaneous reporting systems for drug and vaccine safety have a number of major methodologic weaknesses (see also Chapters 10 and 11). Under-, biased, and incomplete reporting are inherent to all such spontaneous reporting systems and potential safety concerns may be missed.^{118,137} Aseptic meningitis associated with the Urabe mumps vaccine strain, for example, was not detected by spontaneous reporting systems in most countries.^{34,40} Most importantly, however, such spontaneous reports represent less than one-fourth of the information necessary to complete an epidemiologic analysis of a vaccine adverse event.³

Use of data from spontaneous reporting systems are further complicated by lack of specific clinical syndromes being evaluated, absence of laboratory confirmation of many of the events, and simultaneous vaccinations which make determination of which vaccine might have caused the event difficult.

Current spontaneous reporting systems are also prone to detecting increases in adverse events that are not true increases. Instead, they may be due to an increase in (i) reporting efficiency, (ii) vaccine coverage, or (iii) other causes of the adverse event. Spontaneous reporting systems are usually unable to sort out causally related from coincidentally related adverse events because of inherent methodologic weaknesses. For example, an increase in GBS reports after 1993–94 influenza vaccination was found to be due to improvements in vaccine coverage and increases in GBS independent of vaccination.⁷¹ An increased reporting rate of an adverse event following one hepatitis B vaccine compared to a second brand was likely due to differential distribution of brands in the public versus private sectors, which have differential VAERS reporting rates (higher in the public sector).¹³⁸

These studies highlight the crude nature of the “signal” generated by VAERS and the difficulty in ascertaining which vaccine safety concerns warrant further investigation. Not only are there problems with reporting efficiency and potentially biased reporting, but also precise denominators for calculating true rates are usually not available.

Instead, crude measures such as doses distributed must often be used as surrogates for doses administered. Due to these difficulties, the requirement for manufacturers to notify FDA whenever they receive increased number of reports has recently been dropped.¹³⁹

Historically, most countries have relied on spontaneous reporting systems alone for post-licensure vaccine safety monitoring. The inadequacy of scientific information on vaccine safety found by the Institute of Medicine is related to these methodologic weaknesses inherent to spontaneous reporting systems. The establishment of new population based immunization registries, in which all vaccines administered are entered, may provide more timely submission of spontaneous reports as well as more accurate and specific denominators for doses administered, providing information necessary to calculate more accurate adverse event rates.^{140–142}

Clinical Trials

Post Licensure Clinical Trials

Immunization programs are in a dynamic relationship with their target diseases.¹⁴³ To optimize vaccine use, clinical trials may be conducted after vaccine licensure to assess the effects of changes in vaccine formulation,¹⁴⁴ vaccine strain,¹⁴⁵ age at vaccination,¹⁴⁶ the number and timing of vaccine doses,¹⁴⁷ simultaneous administration,¹⁴⁸ and interchangeability of vaccines from different manufacturers¹⁴⁹ on vaccine safety and immunogenicity. The importance of such trials was demonstrated when studies showed an unanticipated differential mortality among recipients of high and regular titer measles vaccine in developing countries,⁵⁶ albeit lower than that among unvaccinated children.¹⁵⁰ This finding resulted in a change in recommendations by WHO for the use of such vaccines.¹⁵¹ The development of automated large linked databases (see below) may permit improved ability to monitor the safety of such postlicensure changes in vaccine use without necessarily conducting such clinical trials.

Phase IV Surveillance Studies

To improve the ability to detect adverse events that are not detected during prelicensure trials, most recently licensed vaccines in developed countries have undergone formal phase IV surveillance studies on populations with sample sizes of $\sim 10^5$. These studies have usually used cohorts in health maintenance organizations supplemented by diary or telephone interview. These methods were first extensively used after the licensure of polysaccharide and conjugated Hib vaccines.^{152–154} Postlicensure studies on safety and efficacy of infant DTaP are also continuing.⁵³ Extensive phase IV evaluation of varicella vaccine includes multiyear evaluation for disease incidence, herpes zoster, and a pregnancy registry.^{155,156} Requirements for phase IV evaluation have even been extended to less frequently used vaccines, such as Japanese encephalitis vaccine.¹⁵⁷ A large postlicensure randomized trial for this vaccine was also completed in China recently to improve the available data on its short term safety.¹⁵⁸ Phase IV data, combined with VAERS to signal the association between rotavirus vaccination and intussusception.¹³⁴

Ad Hoc Epidemiologic Studies

Historically, *ad hoc* epidemiologic studies have been employed to assess signals of potential adverse events generated by spontaneous reporting systems, the medical literature, or other mechanisms. Traditional analyses of secular trends (ecologic studies), cohort, and case–control studies have been used to gather information necessary to measure or compare risks of an adverse event following vaccination with risk in the absence of vaccination. Occasionally, data collected for other study outcomes may be reanalyzed to see whether vaccine was causally related or not.^{159,160} Examples of *ad hoc* followup studies to signals of vaccine safety issues are the investigations of poliomyelitis after inactivated²⁶ and oral polio vaccines,¹¹¹ SIDS after DTP vaccination,^{9,73,161–163} encephalopathy after DTP vaccination,^{17,164} meningoencephalitis after mumps vaccination,^{33,65} injection site abscesses postvaccination,¹⁶⁵ GBS after influenza vaccine^{13,70,71} and intussusception after rotavirus vaccine.¹³⁴ Many

such studies have been compiled and reviewed recently by the Institute of Medicine.^{9,10} While automated large linked databases (see below) provide a more cost-effective and flexible framework for hypothesis testing, *ad hoc* epidemiologic studies may still be needed in countries without automated large linked databases, or where the power of the automated large linked databases may be inadequate to answer a question in a timely manner.^{71,113}

Automated Large-Linked Databases

Ad hoc epidemiologic studies of vaccine safety, while potentially informative about vaccine causality, are costly, time consuming, and usually limited to assessment of a single event. As with drug safety research, efforts have increasingly turned to record linkage between automated exposure (immunization records in lieu of pharmacy) files and outcome medical files. The CDC participated in two pilot vaccine safety studies using automated large linked databases in Medicaid and HMO populations, respectively, during the late 1980s.^{166–169} While validating this approach for vaccine safety studies and providing scientifically rigorous results, these studies were limited by their relatively small sample sizes, retrospective design, and focus on the most severe reactions.⁹ These limitations, the constraints of VAERS, and the recognition of the need for improved monitoring of vaccine safety, prompted the CDC to initiate the Vaccine Safety Datalink (VSD) project in 1990.¹¹³ To help overcome the previously identified shortcomings, 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. Selection of staff model prepaid health plans also minimized potential biases for more severe outcomes resulting from data generated from fee-for-service claims.

The VSD has conducted active surveillance on approximately 500 000 children from birth through 6 years of age (75 000 birth cohort, approximately 2% of US population in these age groups) whose parents (or legal guardians) were enrolled in one of four staff model HMOs.¹¹³

Expansion to include all age groups in the study is now complete. Each site encodes their patients' clinical data with unique study identifiers before sending data to CDC annually for merging and analysis, thereby preserving patient confidentiality. Depending on the background incidence of the medical event, the frequency of specific vaccinations, and their relative risk, events attributable to vaccinations as rare as one per 100 000 doses should be detectable within five years.

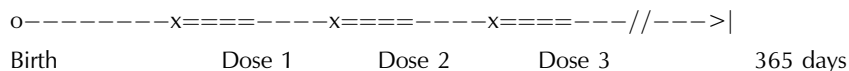
The VSD focused its initial efforts on examining potential associations between immunizations and 34 serious neurologic, allergic, hematologic, infectious, inflammatory, and metabolic conditions. The VSD is also being used to test new *ad hoc* vaccine safety hypotheses. These may arise from the medical literature,¹⁷⁰ from VAERS,¹³⁸ from changes in immunization schedules, or introduction of new vaccines.¹⁷¹ The diversity in vaccination practices at the four HMOs permit useful contrasts in safety experiences.¹⁷¹ The size of the VSD population may also permit

separation of the risks associated with individual vaccines from those associated with vaccine combinations, whether given in the same syringe or simultaneously at different body sites. Such studies will be especially valuable in view of the new combined pediatric vaccines currently in development.¹⁷²

Should the VSD identify an adverse event as being caused by vaccine, data on the incidence rate attributable to vaccine will be available, permitting accurate risk–benefit assessment by both the public and policymakers.¹⁷³ Subgroup analyses may permit identification of risk factors for adverse events, which may be useful in identifying contraindications to vaccinations. Data from VSD should be useful in calculating background rates of illnesses in the absence of vaccination that can serve as expected rates when comparing rates of vaccine associated events in SRS. Also, incidence rates of vaccine associated adverse event derived from VSD can be used to evaluate the sensitivity of passive reporting systems. The VSD data can also

Table 40.4. Example of method for risk interval analysis of association between a universally recommended three-dose vaccine (with few unvaccinated persons for comparison) and adverse event

1. Define "risk interval" for adverse event after vaccination (e.g., 30 days after each dose).
2. Partition observation time for each child in the study into periods within and outside of risk intervals, and sum respectively (e.g., for a child observed for 365 days during which three doses of vaccine were received, total risk interval time = 3×30 person days = 90 person days; total nonrisk interval time = $365 - 90 = 275$ person days).



3. Add up (i) total risk interval and non-risk interval observation times for each child in the study (= person time observed; for mathematical convenience, the example below uses 100 and 1000 person months of observation), and (ii) adverse events occurring in each time period to complete a 2×2 table (for illustration, the example below uses three and 10 cases):

	Adverse event yes	Person time observed (months)	Incidence rate
Vaccinated in risk interval yes	3	100	0.03
Vaccinated in risk interval no	10	1000	0.01
Total	13	1100	

Incidence rate adverse event_{vaccinated} = $3/100 = 0.03$

Incidence rate adverse event_{unvaccinated} = $10/1000 = 0.01$

Relative risk vaccinated : unvaccinated = $0.03/0.01 = 3.0$

Probability finding due to chance: $<5/100$

Conclusion: there is a threefold increase in risk for developing the adverse event within the interval following vaccination

aid the FDA in their evaluation of VAERS data¹³⁸ and the Vaccine Injury Compensation Program in determinations of what events should be compensated as vaccine “injuries”.³⁶

Amid these promises, a few caveats are appropriate. While diverse, the population in the four HMOs currently in the VSD is not wholly representative of the US in terms of geography or socioeconomic status. More importantly, due to the high coverage attained in the HMOs for most vaccines, few nonvaccinated controls are available. The VSD must therefore rely predominantly on some type of “risk interval” analysis^{166–169} (Table 40.4). The capability of this approach to assess associations between vaccination and adverse events with delayed or insidious onset (e.g., autism) is limited. The VSD also cannot easily assess adverse events not currently captured in existing HMO databases, either because they do not result in a healthcare consultation (e.g., fever) or because the data are not automated (e.g., x-ray results).¹¹³ The current VSD is also not large enough to examine the risk of extremely rare events such as Guillain–Barré syndrome after each season’s influenza vaccine. Finally, because the VSD relies on epidemiologic methods, it may not successfully control for confounding and bias in each analysis⁷⁶ and inferences on causality may be limited.¹⁷⁴

Despite these potential shortcomings, the VSD provides a new, essential, powerful, and cost-effective complement to our ongoing evaluations of vaccine safety in the US.¹⁷¹ In view of the methodologic and logistical advantages offered by automated large linked databases, the UK and Canada have also developed automated large linked databases linking immunization registries with medical files.^{40,175} Because of the relatively limited number of vaccines used worldwide and the costs associated with establishing and operating them, it is unlikely that all countries will be able to or need to establish their own automated large linked databases, however. They should be able to draw upon the scientific base established by the existing automated large linked databases for vaccine safety and, if the need arises, conduct *ad hoc* epidemiologic studies.

METHODOLOGIC APPROACHES

Exposures

In countries where vaccinations are required for entry into daycare, kindergarten, schools, and/or colleges, documentation (e.g., vaccination cards or medical records) is usually available and of good quality for most infants and children. In the US, documentation of the vaccine type, date of vaccination, manufacturer, lot number, and vaccine provider in a permanent medical record has been required since 1988 for certain routine childhood vaccinations.⁹¹ This requirement, along with improvements in technology, has prompted many organizations to automate their vaccination records.^{140–142}

Although vaccination records can be manually retrieved and reviewed for any study design, automated vaccination records greatly ease the logistics of organizing such studies. Whenever sampling is necessary in the design, automated records also ease the selection of samples that are representative. Assessing the accuracy of such automated data is important in any study.^{176,177} When persons receive their vaccinations from a variety of providers (as they commonly do in the US), their exposure status may be misclassified.¹⁷⁸ This error could be minimized if a centralized National Vaccination Registry were implemented to track all vaccinations from birth. Such a registry has been implemented in most of the UK¹⁴⁰ and regional registries are under development in the US.¹⁴²

The availability and quality of vaccination records generally decrease as people age. Some vaccines for older people (e.g., tetanus–diphtheria boosters in emergency rooms, hepatitis B vaccinations for healthcare workers) may be administered in settings other than primary healthcare. In addition to review of primary medical records, interviews or a review of data from secondary vaccination sites may therefore be necessary to accurately ascertain exposure status in adverse event studies of these vaccines in older populations. To increase the accuracy of exposure data in a study of adverse reactions to plasma derived hepatitis B vaccine among Alaskan natives,¹⁷⁹ medical records

from the village, the hospital, and the regional public health nurse, in addition to the automated vaccination record, were reviewed.¹⁷⁹ Recent studies of GBS and influenza vaccine relied on both patient/family interview and validation with health provider for exposure ascertainment.^{70,71} Interestingly, reliance on provider verification may lead to underascertainment of vaccination status, either due to poor record keeping¹⁷⁸ or to concerns about liability in vaccine safety studies.⁷¹

To further improve the accuracy and efficiency of transfer of vaccine identification information from the vaccine vial to either automated or paper immunization records, the CDC has organized the Vaccine Identification Standards Initiative (VISI).⁶⁷ Working with a coalition of industry, immunization providers, and other federal agencies, the initiative seeks to develop standards for (i) abbreviations for new vaccine antigens and vaccine manufacturers, (ii) peel-off labels, (iii) barcodes, (iv) lot numbers, (v) immunization records, and (vi) presentation of key identifier information on vaccine packaging (*à la* nutrition label). All of these should contribute to minimizing exposure misclassification for any vaccine related study.

Outcomes

To ensure both high sensitivity and specificity, a sequential approach is usually required. The initial screening definition is highly sensitive but less specific. For a study of neurologic illness following DTP immunization, a combination of hospitalization codes and prescriptions of any drug that might have been used to treat such illnesses were used for case finding.¹⁶⁷ The medical records of these patients are then reviewed (sometimes after chart abstraction) to see if they meet the study case definition. In difficult diagnoses such as GBS, a panel of specialists may also be asked to review the charts after exposure status has been masked.^{22,70}

Should the concern be a new previously undescribed syndrome, analyses of existing databases may be inadequate. A recent study of "Gulf War syndrome" and vaccinations relied on a thorough interview of patients meeting a *de novo* complex case definition before linkage with vaccination history.¹⁸⁰

Analyses

Different analytical strategies are needed depending on how a vaccine is used in the population. For low frequency use vaccines for which the vaccinees are generally no different than the nonvaccinees (e.g., travel vaccines), comparison between the two groups with adequate matching or adjustment is relatively straightforward. For vaccines that are almost universally recommended (e.g., most childhood vaccines), however, too few persons are "unexposed" to compare adverse outcomes among vaccinated and unvaccinated (i.e., typical cohort study) or vaccination among cases and noncases (i.e., typical case-control study). Therefore, an alternative definition of "exposure" is used in most postlicensure vaccine safety studies. A "window of risk" after vaccination for the specific event of interest is defined *a priori* based on current understanding of the most plausible biologic mechanism should such an association actually exist. Incidence rates inside ("exposed") and outside ("unexposed") the risk window (or recent exposures among cases and noncases) are then compared. To allow for the possibility of delayed reactions, multiple risk windows may be defined (e.g., 1, 2–3, 4–7, 8–14, 15–30, 31 or more days; or even longer for autoimmune demyelinating diseases such as GBS), preferably *a priori*. Several calculations typically should be performed by using different intervals, or windows, to identify the window in which the rate ratio or other estimate of risk is greatest. For evaluating hypothesized causes and for establishing new hypotheses, this information is invaluable.

Conditional logistic regression, which should be used for matched case-control studies, can also be used in case series studies.¹⁸¹ Unlike conventional analyses of studies involving matching, to which only discordant pairs contribute, all exposed cases contribute to case series analyses, increasing their potential efficiency. Because cases serve as their own controls, moreover, this design controls within-individual variation perfectly.

Most members of populations under active surveillance contribute little to analysis because few experience adverse events, but the *per capita* cost of information is more or less constant. Thus,

investigators sometimes use the case-cohort method, which requires information only about the cases and samples of the population to which they belong, the latter of which can be used repeatedly with different outcomes (i.e., fewer people must be studied). Efficiency is even further increased for the case crossover method (i.e., there are no controls). Both of these designs are discussed in more detail in Chapter 44).

Confounding and Bias

Once a universally recommended vaccine is licensed, it is usually unethical to withhold the vaccine in subsequent randomized trials as a means to minimize confounding and bias. If further assessment of short term safety is critical, however, it may be possible to do a randomized trial postlicensure in certain settings, as long the controls also receive the vaccine after a relatively short observation period.¹⁵⁸

To minimize recall bias, it is best to rely on data sources that gather information on outcomes and vaccine exposure independently. This is one of the major advantages of record linkage studies.^{113,182} In a recent study of GBS and influenza vaccine, for example, all GBS cases meeting predefined case definition were ascertained from statewide hospital discharge datasets.⁷¹ The cases were then interviewed for a range of risk factors for GBS, including vaccination. Only when patients reported that they were vaccinated were they asked which type. The patients are unaware of the specific hypothesis being evaluated (influenza vaccine and GBS) throughout the interview. A separate random digit dialing survey was conducted to ascertain the comparison influenza vaccination rate in the general population.

Control for potential confounders may be addressed routinely by obtaining the necessary information and adjusting for them at the time of analysis. For the CDC Vaccine Safety Datalink project, sources of covariates include birth certificates and the decennial census (linked via postal zip codes of the children). Alternatively, the case series design¹⁸¹ or assessment of unusual clustering in onset intervals may provide additional information on possible causal association with minimal

confounding. The case series design focuses on changes in disease events through time within individuals, with measured changes pooled over individuals. It has nearly as much statistical power as the full cohort approach if a high percentage of the population has been vaccinated. Since only known vaccinated cases are used in the analysis, there is no exposure misclassification bias. On the negative side, the case series design is still subject to confounding by “within-person” variables that change through time. For example, no epidemiologic method yet can address the “healthy child” bias resulting from avoiding vaccinations during transient illness. Also, research questions may not be limited to events that are time varying (e.g., gender, race/ethnicity).¹⁸²

THE FUTURE

Many persons look to vaccines as the “magic bullet” solution to a number of public health problems that range from acquired immunodeficiency syndrome (AIDS) to malaria. Rapid advances in biotechnology have brought the promise of these new vaccines closer to reality.¹⁸³ Novel delivery technologies such as DNA vaccines and new adjuvants are being explored to permit more antigens to be combined, reducing the number of injections.¹⁷² These changes in vaccines and vaccine delivery will continue to provide additional challenges in proving their safety to an increasingly skeptical and risk averse public, however. Combined with methodologic difficulties associated with studying rare, delayed, or insidious vaccine safety allegations, well organized anti-vaccine organizations, media eagerness for controversy,¹⁸⁴ and relatively rare individual encounters with VPD, vaccine safety concerns are unlikely “to go away” in mature immunization programs.¹⁸⁵

Concomitantly, vaccine safety concerns have also “emerged” as an issue in EPI’s developing countries.¹⁰⁰ The EZ measles vaccine mortality experience highlighted the importance of improving the quality control and evaluating the safety of vaccines used in developing countries.^{56,186} Plans to eliminate neonatal tetanus and measles via National Immunization Days where millions of

persons receive parenteral immunizations over a period of days¹⁸⁷ pose substantial challenges to ensuring injection safety.¹⁸⁸

The increasing computerization and centralization of healthcare services may facilitate epidemiologic studies to reassure the public about the safety of future vaccines. Like other arenas concerned with safety (e.g. aviation, food, blood), a comprehensive "systems" design approach to minimize risk and promote safety is needed. Developments in biotechnology may continue to offer better safer vaccines. The availability of computerized immunization registries may permit optimal implementation of immunization policies at the individual level, ensuring receipt of indicated vaccines, avoiding extra vaccinations, and assuring appropriate observance of valid contraindications to vaccinations. On a longer horizon, vaccine safety research combined with genetic epidemiology may permit better characterization of risk groups for vaccine reactions.^{189,190} Integrated with immunization registries for both children and adults, this may ultimately offer the possibility for better prevention of both vaccine preventable and vaccine induced diseases.

REFERENCES

1. World Bank. *World Development Report 1993: Investing in Health*. New York: Oxford University Press, 1993.
2. *Global Programme for Vaccines and Immunization: Program Report 1997*. WHO/GPV/98.01. Geneva: World Health Organization, 1997.
3. Chen R. Safety of vaccines. In: Plotkin SA, Orenstein WA, eds. *Vaccines*. Philadelphia, PA: Saunders, 1999; 1144–63.
4. Gangarosa EJ, Galazka AM, Wolfe CR, Phillips LM, Gangarosa, RE, Miller E, Chen RT. Impact of the anti-whole-cell pertussis vaccine movements. *Lancet* 1998; **351**: 356–61.
5. Orenstein WA. DTP vaccine litigation, 1988. *Am J Dis Child* 1990; **144**: 517.
6. Institute of Medicine. Liability for the production and sale of vaccine. In: *Vaccine Supply and Innovation*. Washington, DC: National Academy Press, 1985; 85–122.
7. CDSC. Sentinel surveillance shows small decline in MMR coverage. *Commun Dis Rep CDR Wkly* 1998; **8**: 317–20.
8. Nau JY. Hépatite B: une vaccination en souffrance. *Le Monde* 2000 Jun 5: 18 (Col. 1).
9. Howson CP, Howe CJ, Fineberg HV, eds. *Adverse Effects of Pertussis and Rubella Vaccines: a Report of the Committee to Review the Adverse Consequences of Pertussis and Rubella Vaccines*. Washington, DC: National Academy Press; 1991.
10. Stratton KR, Howe CJ, Johnston RB, eds. *Adverse Events Associated with Childhood Vaccines: Evidence Bearing on Causality*. Washington, DC: National Academy Press; 1994.
11. Miller E, Waight P, Farrington P. Safety assessment post-licensure. *Dev Biol Stand* 1998; **95**: 235–43.
12. Miller DL, Wadsworth J, Diamond J, Ross E. Pertussis vaccine and whooping cough as risk factors in acute neurologic illness and deaths in young children. *Dev Biol Stand* 1985; **61**: 389–94.
13. Schonberger LB, Bregman DJ, Sullivan-Bolyai JZ, et al. Guillain-Barré syndrome following vaccination in the national influenza immunization program, United States, 1976–1977. *Am J Epidemiol* 1979; **110**: 105–23.
14. Centers for Disease Control and Prevention. Poliomyelitis prevention in the United States: introduction of a sequential vaccination schedule of inactivated poliovirus vaccine followed by oral poliovirus vaccine. Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep* 1997; **46**: 1–25.
15. McGoldrick MD, Bailie GR. Nonnarcotic analgesics: prevalence and estimated economic impact of toxicities. *Ann Pharmacother* 1997; **31**: 221–7.
16. Marcuse EK, Wentz KR. The NCES reconsidered: summary of a 1989 workshop. *Vaccine* 1990; **8**: 531–5.
17. Alderslade R, Bellman MH, Rawson NSB, Ross EM, Miller DL. The National Childhood Encephalopathy Study. In: *Whooping Cough: Reports from the Committee on the Safety of Medicines and the Joint Committee on Vaccination and Immunisation*. Department of Health and Social Security. London: HM Stationery Office, 1981; 79–169.
18. Wentz KR, Marcuse EK. Diphtheria-tetanus-pertussis vaccine and serious neurologic illness: an updated review of the epidemiologic evidence. *Pediatrics* 1991; **87**: 287–97.
19. Neustadt RE, Fineberg HV. *The Swine Flu Affair: Decision-Making on a Slippery Disease*. Washington, DC: US Government Printing Office, 1978.
20. Kurland LT, Wiederholt WC, Kirkpatrick JW, Potter HG, Armstrong P. Swine influenza vaccine and Guillain-Barré syndrome. Epidemic or artifact? *Arch Neurol* 1985; **42**: 1089–90.

21. Langmuir AD, Bregman DJ, Kurland LT, Nathanson N, Victor M. An epidemiologic and clinical evaluation of Guillain-Barré syndrome reported in association with the administration of swine influenza vaccines. *Am J Epidemiol* 1984; **119**: 841-79.
22. Safranek TJ, Lawrence DN, Kurland LT, Culver DH, Wiederholt WC, Hayner NS, *et al.* Reassessment of the association between Guillain-Barré syndrome and the receipt of swine influenza vaccine in 1976-1977: results of a two-state study. *Am J Epidemiol* 1991; **133**: 940-51.
23. Schumacher W. Legal/ethical aspects of vaccinations. *Dev Biol Stand* 1979; **43**: 435-8.
24. Hardin G. The tragedy of the commons. *Science* 1968; **162**: 1243-8.
25. Fine PEM, Clarkson JA. Individual versus public priorities in the determination of optimal vaccination policies. *Am J Epidemiol* 1986; **124**: 1012-20.
26. Nathanson N, Langmuir AD. The Cutter incident. *Am J Hygiene* 1963; **78**: 16-81.
27. Wilson GS. *The Hazards of Immunization*. London: Athlone, 1967.
28. Butel JS, *et al.* Cell and molecular biology of simian virus 40: implications for human infections and disease. *J Natl Cancer Inst* 1999; **91**: 119-34.
29. Food and Drug Administration. Bovine-derived materials: agency letters to manufacturers of FDA-regulated products. *Fed Register* 1994; **59**: 44591-4.
30. Parkman PD, Hardegree MC. Regulation and testing of vaccines. In: Plotkin SA, Orenstein WA, eds, *Vaccines*. Philadelphia, PA: Saunders, 1999; 1100-43.
31. Anonymous. Two MMR vaccines withdrawn. *Lancet* 1992; **340**: 922.
32. Schmitt HJ, Just M, Neiss A. Withdrawal of a mumps vaccine: reasons and impacts. *Eur J Pediatr* 1993; **152**: 387-8.
33. Kimura M, Kuno-Sakai H, Yamazaki S, *et al.* Adverse events associated with MMR vaccines in Japan. *Acta Paediatr Japan* 1996; **38**: 205-11.
34. Lloyd JC, Chen RT. The Urabe mumps vaccine: lessons in adverse event surveillance and response. *Pharmacoepidemiol Drug Safety* 1996; **5**: S45.
35. Evans G. Vaccine injury compensation programs worldwide. *Vaccine* 1999; **17** (Suppl 3): S25-35.
36. Evans G. Vaccine liability and safety: a progress report. *Pediatr Infect Dis J* 1996; **15**: 477-8.
37. Kitch EW, Evans G. U.S. Law. In: Plotkin SA, Orenstein WA, eds, *Vaccines*. Philadelphia, PA: Saunders, 1999; 1164-99.
38. Centers for Disease Control. Public Health Service recommendations on smallpox vaccination. *MMWR Morb Mortal Wkly Rep* 1971; **20**: 339-45.
39. Bottiger M. The elimination of polio in the Scandinavian countries. *Public Health Rev* 1993-94; **21**: 27-33.
40. Farrington CP, Pugh S, Colville A, *et al.* A new method for active surveillance of adverse events from diphtheria/tetanus/pertussis and measles/mumps/rubella vaccines. *Lancet* 1995; **345**: 567-9.
41. Stetler HC, Orenstein WA, Bart KJ, Brink EW, Brennan JP, Hinman AR. History of convulsions and use of pertussis vaccine. *J Pediatr* 1985; **107**: 175-9.
42. Robbins JB, Pittman M, Trollfors B, Lagergard TA, Taranger J, Schneerson R. Primum non nocere: a pharmacologically inert pertussis toxoid alone should be the next pertussis vaccine. *Pediatr Infect Dis J* 1993; **12**: 795-807.
43. Church JA, Richards W. Recurrent abscess formation following DTP immunizations: association with hypersensitivity to tetanus toxoid. *Pediatrics* 1985; **75**: 899-900.
44. Brown EG, Dimock K, Wright KE. The Urabe AM9 mumps vaccine is a mixture of viruses differing at amino acid 335 of the hemagglutinin-neuraminidase gene with one form associated with disease. *J Infect Dis* 1996; **174**: 619-22.
45. Kew OM, Nottay BK. Molecular epidemiology of polioviruses. *Rev Infect Dis* 1984; **6**: S499-S504.
46. Plotkin SA, Koprowski H. Rabies vaccine. In: Plotkin SA, Mortimer EA, eds, *Vaccines*. Philadelphia, PA: Saunders, 1994: 649-70.
47. Rosenthal KL, McVittie LD. The clinical testing of preventive vaccines. In: Mathieu M, ed., *Biologic Development: a Regulatory Overview*. Waltham, MA: Paraxel, 1993; 119-30.
48. Pinichiero ME. Acellular pertussis vaccines: towards an improved safety profile. *Drug Experience* 1996; **15**: 311-24.
49. Francis T, Kornis RF, Voight RB, *et al.* An evaluation of the 1954 poliomyelitis vaccine trials: summary report. *Am J Public Health* 1955; **45**: 1-63.
50. Eskola J, Kayhty H, Takala AK, *et al.* A randomized, prospective field trial of a conjugate vaccine in the protection of infants and young children against invasive *Haemophilus influenzae* type b disease. *New Engl J Med* 1990; **323**: 1381-7.
51. Black SB, Cherry JD, Shinefield HR, *et al.* Apparent decreased risk of invasive bacterial disease after heterologous childhood immunization. *Am J Dis Child* 1991; **145**: 746-9.
52. Chen RT. Vaccine risks: real, perceived, and unknown. *Vaccine* 1999; **17** (Suppl 3): S41-6.
53. Brown F, Greco D, Mastrantonio P, Salmaso S, Wassilak S. *Pertussis Vaccine Trials*. Basel: Karger, 1997.
54. Heijbel H, Atti MC, Harzer E, *et al.* Hypotonic hyporesponsive episodes in eight pertussis vaccine trials. *Dev Biol Stand* 1997; **89**: 101-3.

55. Hall AJ, Cutts FT. Lessons from measles vaccination in developing countries. *Br Med J* 1993; **307**: 1294–5.
56. Knudsen KM, Aaby P, Whittle H, Rowe M, Samb B, Simondon F, Stevne J, Fine P. Child mortality following standard, medium, or high titre measles immunization in West Africa. *Int J Epidemiol* 1996; **25**: 665–73.
57. Ray WA, Griffin MR. Confounding in studies of adverse reactions to vaccines. *Am J Epidemiol* 1994; **139**: 229.
58. Fine PEM. Methodological issues in the evaluation and monitoring of vaccine safety. *Ann NY Acad Sci* 1995; **754**: 300–8.
59. Baraff LJ, Cody CL, Cherry JD. DTP-associated reactions: an analysis by injection site, manufacturer, prior reactions, and dose. *Pediatrics* 1984; **73**: 31–6.
60. Baraff LJ, Manclark CR, Cherry JD, Christenson P, Marcy SM. Analyses of adverse reactions to diphtheria and tetanus toxoids and pertussis vaccine by vaccine lot, endotoxin content, pertussis vaccine potency and percentage of mouse weight gain. *Pediatr Infect Dis J* 1989; **8**: 502–7.
61. Division of Immunization. Vaccine-associated adverse events in Canada, 1992 report. *Can Commun Dis Rep* 1995; **21**: 117–28.
62. Chen RT, Rastogi SC, Mullen JR, *et al.* The Vaccine Adverse Event Reporting System (VAERS). *Vaccine*. 1994; **12**: 542–50.
63. Ellenborg SS, Chen RT. The complicated task of monitoring vaccine safety. *Public Health Rep* 1997; **112**: 10–20.
64. Wise RP, Kiminyo KP, Salive ME. Hair loss after routine immunizations. *J Am Med Assoc* 1997; **278**: 1176–8.
65. Miller E, Goldacre M, Pugh S, Colville A, Farrington P, Flower A, *et al.* Risk of aseptic meningitis after measles, mumps, and rubella vaccine in UK children. *Lancet* 1993; **341**: 979–82.
66. Edsall G, Elliot MW, Peebles TC *et al.* Excessive use of tetanus toxoid boosters. *J Am Med Assoc* 1967; **202**: 111–3.
67. The Vaccine Identification Standards Initiative. Atlanta, GA, US Centers for Disease Control and Prevention, National Immunization Program 2000 (Internet communication June 20, 2000 at <http://www.cdc.gov/nip/visi>).
68. Halsey NA. Practical considerations regarding the impact on immunization schedules of the introduction of new combined vaccines. *Ann NY Acad Sci* 1995; **754**: 250–4.
69. Institute of Medicine. *Report of Committee to Review an Epidemiologic Study of Neurologic Illness and Vaccination in Children*. Washington, DC: Institute of Medicine, 1990.
70. Chen RT, Kent JH, Rhodes PH, Simon P, Schonberger LS. Investigation of a possible association between influenza vaccination and Guillain–Barré syndrome in the United States, 1990–91. *Postmarketing Surveillance* 1992; **6**: 5–6.
71. Lasky T, Terracciano GJ, Magder L, *et al.* The Guillain–Barré Syndrome and the 1992–93 and 1993–94 influenza vaccines. *New Engl J Med* 1998; **339**: 1797–802.
72. Laboratory Centre for Disease Control. Report of the working group on the possible relationship between hepatitis B vaccination and the chronic fatigue syndrome. *Can Med Assoc J* 1993; **149**: 314–9.
73. Bernier RH, Frank JA, Dondero TJ, Turner P. Diphtheria–tetanus–toxoids–pertussis vaccination and sudden infant deaths in Tennessee. *J Pediatr* 1982; **101**: 419–21.
74. Wakefield AJ, Murch SH, Anthony A, *et al.* Ileal lymphoid nodular hyperplasia, non-specific colitis, and regressive developmental disorder in children. *Lancet* 1998; **351**: 637–41.
75. Mortimer EA, Jones PK, Adelson L. DTP and SIDS. *Pediatr Infect Dis J* 1983; **2**: 492–3.
76. Fine PEM, Chen RT. Confounding in studies of adverse reactions to vaccines. *Am J Epidemiol* 1992; **136**: 121–35.
77. Rastogi SC, ed. *International Workshop: Harmonization of Reporting of Adverse Events Following Vaccination*. Rockville MD: FDA/CBER, 1993.
78. Wiholm BE, Olsson S, Moore N, Wood S. Spontaneous reporting systems outside the U.S. In: Strom BL, ed., *Pharmacoepidemiology*. Chichester: Wiley, 1994; 139–55.
79. Jonville-Bera AP, Autret E, Galy-Eyraud C, Hessel L. Thrombocytopenic purpura after measles, mumps and rubella vaccination. *Pediatr Infect Dis J* 1996; **5**: 44–8.
80. Mansoor O, Pillans PI. Vaccine adverse events reported in New Zealand 1990–5. *NZ Med J* 1997; **110**: 270–2.
81. Taranger J, Holmberg K. Urgent to introduce countrywide and systematic evaluation of vaccine side effects. *Lakartidningen* 1992; **89**: 1691–3.
82. Salisbury DM, Beggs NT, eds. *Immunisation Against Infectious Disease*. London: HMSO, 1996; 29–33.
83. Mehta U, Milstien JB, Ducbs P, Folb PI. Developing a national system for dealing with adverse events following immunization. *Bull WHO* 2000; **78**: 170–7.
84. Anonymous. Surveillance of serious adverse events following vaccination. *Commun Dis Intell* 1995; **19**: 273–4.
85. Duclos P. Surveillance des effets secondaires de la vaccination. *Sante* 1994; **4**: 215–20.

86. Andersen MM, Ronne T. Side-effects with Japanese encephalitis vaccine. *Lancet* 1991; **337**: 1044.
87. Sokhey, J. Adverse events following immunizations: 1990. *Indian Pediatr* 1991; **28**: 593–607.
88. Squarcione S, Vellucci L. Adverse reactions following immunization in Italy in the years 1991–93. *Ig Mod* 1996; **105**: 1419–31.
89. Health Council of the Netherlands, Committee on Adverse Reactions to Vaccinations. *Adverse Reactions to Vaccines in the National Immunization Programme in 1993*, publication 1995/08E. The Hague: Health Council of the Netherlands, 1995.
90. Brito GS. *System of Investigation and Notification of Adverse Events Following Immunization, Preliminary Report*. Sao Paulo: Health Department, 1991.
91. Centers for Disease Control. National Childhood Vaccine Injury Act: requirements for permanent vaccination records and the reporting of selected events after vaccination. *MMWR Morb Mortal Wkly Rep* 1988; **37**: 197–200.
92. Stetler HC, Mullen JR, Brennan JP, Livengood JR, Orenstein WA, Hinman AR. Monitoring system for adverse events following immunization. *Vaccine* 1987; **5**: 169–174.
93. Food and Drug Administration (FDA). *COST-ART—Coding Symbols for Therapeutic Adverse Reaction Terms*, 3rd edn. Rockville, MD: FDA, 1989.
94. Braun MM, Ellenberg SS. Descriptive epidemiology of adverse events following immunization: reports to the Vaccine Adverse Event Reporting System (VAERS), 1991–1994. *J Pediatr* 1997; **131**: 529–35.
95. Morris R, Halperin S, Dery P, *et al.* IMPACT monitoring network: a better mousetrap. *Can J Infect Dis* 1993; **4**: 194–5.
96. Collet JP, MacDonald N, Cashman N, Pless R and the Advisory Committee on Causality Assessment. Monitoring signals for vaccine safety: the assessment of individual adverse event reports by an expert advisory committee. *Bull WHO* 2000; **78**: 178–85.
97. Edwards IR, Biriell C. Harmonisation in pharmacovigilance. *Drug Safety* 1994; **10**: 93–102.
98. Food and Drug Administration, International Conference on Harmonisation. Guideline on Clinical Safety Data Management: periodic safety update reports for marketed drugs; availability. *Fed Register* 1997; **62**: 27 470–6.
99. World Health Organization (WHO). *Surveillance of Adverse Events Following Immunization*, WHO/EPI/TRAM/93.02 Rev. 1. Geneva: WHO, 1997.
100. World Health Organization (WHO). Vaccine supply and quality: surveillance of adverse events following immunization. *Wkly Epidemiol Rec* 1996; **71**: 237–42.
101. Global Programme on Vaccines and Immunization. *Quality: The First Consideration*, WHO/GPV/SAGE.97/WP.03. Geneva: WHO, 1997.
102. Fenichel GM, Lane DA, Livengood JR, Horwitz SJ, Menkes JH, Schwartz JF. Adverse events following immunization: assessing probability of causation. *Pediatr Neurol* 1989; **5**: 287–90.
103. Wassilak SGF, Sokhey J. *Monitoring of Adverse Events Following Immunization in the Expanded Programme on Immunization*, WHO/EPI/GEN/91.2. Geneva: WHO, 1991; 1–29.
104. Goodman M, Lamm SH, Bellman MH. Temporal relationship modeling: DTP or DT immunizations and infantile spasms. *Vaccine* 1998; **16**: 225–31.
105. Laboratory Centre for Disease Control. Proceedings of a workshop on the standardization of definitions for post-marketing surveillance of adverse vaccine reactions, Ottawa, 1992. Ottawa: LCDC, 1992; 1–17.
106. Kramer MS, Lane DA. Causal propositions in clinical research and practice. *J Clin Epidemiol* 1992; **45**: 639–49.
107. Hill AB. The environment and disease: association or causation. *Proc R Soc Med* 1965; **58**: 295–300.
108. Kramer MS. Difficulties in assessing the adverse effects of drugs. *Br J Clin Pharmacol* 1981; **11**: S105–10.
109. Jones JK. Determining causation from case reports. In: Strom BL, ed., *Pharmacoepidemiology*. Chichester: Wiley, 1994; 365–78.
110. Forsey T, Mawn JA, Yates PJ, Bently ML, Minor PD. Differentiation of vaccine and wild mumps viruses using the polymerase chain reaction and dideoxynucleotide. *J Gen Virol* 1990; **71**: 987–90.
111. Henderson DA, Witte JJ, Morris L, Langmuir AD. Paralytic disease associated with oral polio vaccines. *J Am Med Assoc* 1964; **190**: 153–60.
112. Centers for Disease Control and Prevention. Rubella vaccination during pregnancy: United States, 1971–1988. *MMWR Morb Mortal Wkly Rep* 1989; **38**: 289–93.
113. Chen RT, Glasser J, Rhodes P, *et al.* The Vaccine Safety Datalink Project: a new tool for improving vaccine safety monitoring in the United States. *Pediatrics* 1997; **99**: 765–73.
114. Cherry JD. Pertussis and the vaccine controversy. In: Root RK, *et al.*, eds, *Contemporary Issues in Infectious Diseases*. London: Churchill Livingstone, 1989; 47–63.
115. Lane DA. A probabilist's view of causality assessment. *Drug Inf J* 1984; **18**: 323–30.
116. Venulet J, ed. *Assessing Causes of Adverse Drug Reactions with Special Reference to Standardized Methods*. London: Academic, 1982.

117. Lane DA, Kramer MS, Hutchinson TA, Jones JK, Naranjo C. The causality assessment of adverse drug reactions using a Bayesian approach. *Pharmaceut Med* 1987; **2**: 265–83.
118. Finney DJ. The detection of adverse reactions to therapeutic drugs. *Stat Med* 1982; **1**: 153–61.
119. Retailiau HF, Curtis AC, Storr G, Caesar G, Eddins DL, Hattwick MAW. Illness after influenza vaccination reported through a nationwide surveillance system, 1976–77. *Am J Epidemiol* 1980; **111**: 270–8.
120. Bate A, Lindquist M, Edwards IR, et al. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol* 1998; **54**: 315–21.
121. Uhari M, Rantala H, Niemelä M. Cluster of childhood Guillain–Barré cases after an oral poliovaccine campaign. *Lancet* 1989; **2**: 440–1.
122. Berg SW, Mitchell BS, Hanson RK, et al. Systemic reactions in U.S. Marine Corps personnel who received Japanese encephalitis vaccine. *Clin Infect Dis* 1997; **24**: 265–6.
123. Kilroy AW, Schaffner W, Fleet WF Jr, et al. Two syndromes following rubella immunization. Clinical observations and epidemiological studies. *J Am Med Assoc* 1970; **214**: 2287–92.
124. Yergeau A, Alain L, Pless R, Robert Y. Adverse events temporally associated with meningococcal vaccines. *Can Med Assoc J* 1996; **1154**: 503–7.
125. Committee on Safety of Medicines and Medicines Control Agency. Adverse reactions to measles–rubella vaccine. *Current Problems Pharmacovigilance* 1995; **2**: 9–10.
126. da Silveira CM, Salisbury DM, de Quadros CA. Measles vaccination and Guillain–Barré syndrome. *Lancet* 1997; **349**: 14–6.
127. Rantala H, Cherry JD, Shields WD. Epidemiology of Guillain–Barré syndrome in children: relationship of oral polio vaccine administration to occurrence. *J Pediatr* 1994; **124**: 220–3.
128. Carson JL, Strom BL, Maislin G. Screening for unknown effects of newly marketed drugs. In: Strom BL, ed., *Pharmacoepidemiology*. Chichester: Wiley, 1994; 431–47.
129. Chen RT, Wattigney W, Garguillo P. Proactive surveillance in vaccine safety. *Pharmacoepidemiol Drug Safety* 2000; in press.
130. Singleton, JA, Lloyd JC, Terracciano GJ, Salive ME, Chen RT, VAERS. An overview of the Vaccine Adverse Event Reporting System as a surveillance system. *Vaccine* 1999; **17**: 2908–17.
131. Beeler J, Varricchio F, Wise R. Thrombocytopenia after immunization with measles vaccines: review of the vaccine adverse events reporting system (1990 to 1994). *Pediatr Infect Dis J* 1996; **15**: 88–90.
132. Pool V, Glasser J, Garguillo P, Chen RT, and the Vaccine Safety Datalink Team. Risk of allergic reactions to gelatin-containing vaccines: a case–control analysis of the Vaccine Safety Datalink. Oral presentation at the first annual conference on vaccine research: *Basic Science—Product Development—Vaccine Safety and Adverse Effects*, Washington, DC, 1998.
133. Braun MM, Patriarca PA, Ellenberg SS. Syncope after immunization. *Arch Pediatr Adolesc Med* 1997; **151**: 255–9.
134. CDC. Intussusception among recipients of rotavirus vaccine—United States, 1998–1999. *MMWR Morb Mortal Wkly Rep* 1999; **48**: 577–81.
135. Rosenthal S, Chen R, Hadler SC. The safety of acellular pertussis vaccine versus whole cell pertussis vaccine: a post-marketing assessment. *Arch Pediatr Adolesc Med* 1996; **150**: 457–60.
136. Rastogi SC, Wise RP. Exploring associations between vaccine lot potencies and human safety experience. *Pharmacoepidemiol Drug Safety* 1996; **5**: S75.
137. Rosenthal S, Chen RT. Reporting sensitivities of two passive surveillance systems for vaccine adverse events. *Am J Public Health* 1995; **85**: 1706–9.
138. Niu MT, Rhodes P, Salive M, et al. Comparative safety of two recombinant hepatitis b vaccines in children: data from the Vaccine Adverse Event Reporting System (VAERS) and Vaccine Safety Datalink (VSD). *J Clin Epidemiol* 1998; **51**: 503–10.
139. Food and Drug Administration. Postmarketing expedited adverse experience reporting for human drug and licensed biological products; increased frequency reports. *Fed Register* 1997; **62**: 34166–8.
140. Begg NT, Gill ON, White JM. COVER (cover of vaccination evaluated rapidly): description of the England and Wales scheme. *Public Health* 1989; **103**: 81–9.
141. Cordero JF, Guerra FA, Saarlans KN, eds. Developing national immunization registries: experience from the All Kids Count Program. *Am J Prev Med* 1997; **13** (suppl 1): 1–128.
142. Wood D, Saarlans KN, Inkelas M, Matyas BT. Immunization registries in the United States: implications for the practice of public health in a changing health care system. *Annu Rev Public Health* 1999; **20**: 231–55.
143. Chen RT, Orenstein WA. Epidemiologic methods in immunization programs. *Epidemiol Rev* 1996; **18**: 99–117.
144. Patriarca PA, Laender F, Palmeira G, et al. Randomised trial of alternative formulations of oral poliovaccine in Brazil. *Lancet* 1988; **1**: 429–34.
145. Bhargava I, Chharparwal BC, Phadke MA, et al. Reactogenicity of indigenously produced measles vaccine. *Indian Pediatr* 1996; **33**: 827–31.
146. Orenstein WA, Markowitz LE, Preblud SR, et al. Appropriate age for measles vaccination in the United States. *Dev Biol Stand* 1986; **65**: 13–21.

147. Booy R, Aitken SJM, Taylor S, *et al.* Immunogenicity of combined diphtheria, tetanus, and pertussis vaccine given at 2, 3, and 4 months versus, 3, 5, and 9 months of age. *Lancet* 1992; **339**: 507–10.
148. Deforest A, Long SS, Lischner HW, *et al.* Simultaneous administration of measles–mumps–rubella vaccine with booster doses of diphtheria–tetanus–pertussis and poliovirus vaccines. *Pediatrics* 1988; **81**: 237–46.
149. Scheifele D, Law B, Mitchell L, *et al.* Study of booster doses of two Haemophilus influenzae type b conjugate vaccines including their interchangeability. *Vaccine* 1996; **14**: 1399–406.
150. Aaby P, Samb B, Simondon F, *et al.* A comparison of vaccine efficacy and mortality during routine use of high titer EZ and Schwarz standard measles vaccines in rural Senegal. *Trans R Soc Trop Med Hyg* 1996; **90**: 326–30.
151. Expanded Programme on Immunization: Safety of high titer measles vaccines. *Wkly Epidemiol Rec* 1992; **67**: 357–61.
152. Black SB, Shinefield HR, and the Northern California Permanente Medical Care Program Department of Pediatrics Vaccine Study Group. b-CAPSA I Haemophilus influenzae, type b capsular polysaccharide vaccine safety. *Pediatrics* 1987; **79**: 321–25.
153. Meekison W, Hutcheon M, Guasparini R, *et al.* Post-marketing surveillance of adverse events following PROHIBIT vaccine in British Columbia. *Can Med Assoc J* 1989; **141**: 927–9.
154. Vadheim CM, Greenberg DP, Marcy SM, Froeschle J, Ward JI. Safety evaluation of PRP-D Haemophilus influenzae type b conjugate vaccine in children immunized at 18 months of age and older: followup study of 30, 000 children. *Pediatr Infect Dis J* 1990; **9**: 555–61.
155. Centers for Disease Control and Prevention. Unintentional administration of varicella virus vaccine: United States, 1996. *MMWR Morb Mortal Wkly Rep* 1996; **45**: 1017–8.
156. Black S, Shinefield H, Ray P, Lewis E, Hansen J, Schwalbe J, Coplan P, Sharrar R, Guess H. Postmarketing evaluation of the safety and effectiveness of varicella vaccine. *Pediatr Infect Dis J* 1999; **18**: 1041–6.
157. Centers for Disease Control and Prevention: Inactivated Japanese encephalitis virus vaccine. Recommendations of the advisory committee on immunization practices (ACIP). *MMWR Morb Mortal Wkly Rep* 1993; **42**: 10.
158. Liu ZL, Hennessy S, Strom BL, *et al.* Short-term safety of live attenuated Japanese encephalitis vaccine (SA14–14–2): results of a randomized trial with 26, 239 subjects. *J Infect Dis* 1997; **176**: 1366–9.
159. Jones P, Fine P, Piracha S. Crohn's disease and measles. *Lancet* 1997; **349**: 473.
160. Hejbel H, Chen RT, Dahlquist G. Cumulative incidence of childhood-onset IDDM is unaffected by pertussis immunization. *Diabetes Care* 1997; **20**: 173–5.
161. Solberg LK. *DPT Vaccination, Visit to Child Health Center and Sudden Infant Death Syndrome (SIDS): Evaluation of DPT Vaccination*, report to the Oslo Health Council. Bethesda, MD: NIH Library Translation 85–152, 1985.
162. Bouvier-Colle MH, Flahaut A, Messiah A, *et al.* Sudden infant death and immunization: an extensive epidemiologic approach to the problem in France: winter 1986. *Int J Epidemiol* 1989; **18**: 121–6.
163. Mitchell EA, Stewart AV, Clements M, *et al.* Immunisation and sudden infant death syndrome. *Arch Dis Child* 1995; **73**: 498–501.
164. Gale JL, Thapa PB, Wassilak SGF, Bobo JK, Mendelman PM, Foy HM. Risk of serious acute neurological illness after immunization with DTP: report of a case–control study. *J Am Med Assoc* 1994; **271**: 37–41.
165. Simon PA, Chen RT, Elliott JA, Schwartz B. Outbreak of pyogenic abscesses after diphtheria and tetanus toxoids and pertussis vaccination. *Pediatr Infect Dis J* 1993; **12**: 368–71.
166. Walker AM, Jick H, Perera DR, *et al.* Diphtheria–tetanus–pertussis immunization and sudden infant death syndrome. *Am J Public Health*. 1987; **77**: 945–51.
167. Walker AM, Jick H, Perera DG, *et al.* Neurologic events following diphtheria–tetanus–pertussis immunization. *Pediatrics*. 1988; **81**: 345–9.
168. Griffin MR, Ray WA, Livengood JR, *et al.* Risk of sudden infant death syndrome following diphtheria–tetanus–pertussis immunization. *New Engl J Med* 1988; **319**: 618–23.
169. Griffin MR, Ray WA, Mortimer EA, *et al.* Risk of seizures and encephalopathy after immunization with the diphtheria–tetanus–pertussis vaccine. *J Am Med Assoc*. 1990; **263**: 1641–5.
170. Ray P, Black S, Shinefield H, Dillon A, *et al.* Risk of chronic arthropathy following rubella vaccination. *J Am Med Assoc* 1997; **278**: 551–6.
171. Chen RT, DeStefano F, Davis RL, Jackson LA, Thompson RS, Mullooly JP, Black SB, Shinefield HR, Vadheim CM, Ward JI, Marcy SM *et al.* The Vaccine Safety Datalink: immunization research in health maintenance organisations in the USA. *Bull WHO* 2000; **78**: 186–94.
172. Williams JC, Goldenthal KL, Burns DL, Lewis BP Jr, eds. *Combined Vaccines and Simultaneous Administration: Current Issues and Perspective*. New York: New York Academy of Science, 1995.

173. Hinman AR, Orenstein WA. Public health considerations: US. In: Plotkin SA, Mortimer EA, eds, *Vaccine*. Philadelphia, PA: Saunders, 1999.
174. Rothman KJ, ed. *Causal Inference*. Chestnut Hill, MA: Epidemiology Resources, 1988.
175. Roberts JD, Roos LL, Poffenroth LA, et al. Surveillance of vaccine-related adverse events in the first year of life: a Manitoba cohort study. *J Clin Epidemiol* 1996; **49**: 51–8.
176. Griffin MR, Ray WA, Fought RL, Foster MA, Hays A, Schaffner W. Monitoring the safety of childhood immunizations: methods of linking and augmenting computerized data bases for epidemiologic studies. *Am J Prev Med* 1988; **4**: S5–S13.
177. Mullooly J, Drew L, DeStefano F, et al. Quality of HMO vaccination databases used to monitor childhood vaccine safety. *Am J Epidemiol* 1999; **149**: 186–94.
178. Wilton R, Pennisi AJ. Evaluating the accuracy of transcribed computer-stored immunization data. *Pediatrics* 1994; **94**: 902–6.
179. McMahon BJ, Helminiak C, Wainwright RB, Bulkow L, Trimble BA, Wainwright K. Frequency of adverse reactions to hepatitis B vaccine in 43,618 persons. *Am J Med* 1992; **92**: 254–6.
180. Unwin C, Blatchley N, Coker W, et al. Health of UK servicemen who served in Persian Gulf War. *Lancet* 1999; **353**: 169–78.
181. Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol* 1996; **143**: 1165–73.
182. Garguillo PM, Kramarz P, DeStefano F, Chen RT. An application of conditional Poisson regression: a new tool for controlling confounding in pharmacoepidemiology. Presented at the 7th biennial symposium on statistical methods, Atlanta, GA, 1999.
183. Rabinovich NR, McInnes P, Klein DL, Hall BF. Vaccine technologies: view to the future. *Science* 1994; **265**: 1401–4.
184. Freed GL, Katz SL, Clark SJ. Safety of vaccinations. Miss America, the media, and public health. *J Am Med Assoc* 1996; **276**: 1869–72.
185. Ward BJ. Vaccine adverse events in the new millenium: is there reason for concern? *Bull WHO* 2000; **78**: 205–15.
186. Dellepiane N, Griffiths E, Milstren JB. New challenges in assuring vaccine quality. *Bull Who* 2000; **78**: 155–62.
187. Olive JM, Risi JB, DeQuadros CA. National immunization days: experience in Latin America. *J Infect Dis* 1997; **175**: S189–93.
188. Hutin Y, Chen RT. Injection safety: a global challenge. *Bull WHO* 1999; **77**: 787–8.
189. Khoury MJ, Risch N, Kelsey JL. Genetic epidemiology. *Epidemiol Rev* 1997; **19**: 1–185.
190. Pope JE, Stevens A, Howson W, Bell DA. The development of rheumatoid arthritis after recombinant hepatitis B vaccination. *J Rheumatol* 1998; **25**: 1687–93.

41

Special Methodological Issues in Pharmacoepidemiology Studies of Devices*

ROSELIE A. BRIGHT

Epidemiology Branch, Division of Postmarket Surveillance, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, Food and Drug Administration, USA

INTRODUCTION

Early in 1990, the FDA Center for Devices and Radiological Health (CDRH) received reports of two rapid deaths associated with barium enemas (used to provide x-ray contrast media for colon imaging). The interesting feature was that the middle-aged women, in different parts of the country, had suffered their reactions after the barium enema cuffs were put in place, but before barium had been introduced. The reactions were clearly consistent with anaphylaxis, although in one instance the reporter called it a vasovagal reaction. The medical community was already familiar with allergy to barium, and the Center for Drug Evaluation and Research (CDER) already had many such reports on hand. However, in these

two cases, the only potential allergens were the lubricants or the cuffs.

The scope of investigation was widened to look at all reports associated with barium enemas in the files of both centers and the manufacturer of the cuff, which held most of the market. FDA had received, from all sources, reports of five deaths and 28 very serious reactions, all of an allergic nature (except for one death due to aspiration of oral barium). All of the deaths occurred in 1989 or 1990. There was a preponderance of middle-aged female victims.

A literature search found a fatality immediately after and attributed to a single contrast colonic infusion reported in 1989,¹ and a 1990 publication reporting seven reactions among 6918 barium enema infusions at one hospital from January

* The opinions expressed in this chapter are those of the author and do not necessarily represent the official policies of the US Food and Drug Administration.

1987 to March 1989, attributed to some unknown new additive in the barium enema infusion.²

At this point, the evidence implicated the cuff rather than the lubricant. However, there were still many other questions to consider:

- Was the problem specific to the manufacturer, the cuff, or the material?
- Could the problem be mitigated by changing manufacturing procedures?
- If the cuff were to be recalled, would substitute products be available?
- Was this a trend or an isolated, fixable problem?
- What were the public health implications?

During the 1990s, it became apparent that the culprit was latex, which is more or less allergenic depending on the manufacturing process.³ Humans contact latex in many different devices, including condoms, diaphragms, anesthetic gas masks (many reactions attributed to gas may have been to latex, instead), airway tubes, catheter tip occluders, dental dams, tympanostomy tubes, and gloves. Allergenic particles have readier access to immune cells through surgical wounds or mucous membranes than through intact skin, and one can see that many of the listed devices have direct contact with the former.³ Furthermore, the health care profession and patient population had been experiencing greater latex exposure throughout the 1980s with the introduction of universal precautions⁴ and promotion of condom use. The stage had been set for a serious latex allergy problem.

So, where are the numbers? How many people are or were exposed to latex? What is or was the extent of exposure? Does anatomical site of exposure matter in the way that was hypothesized? What are the rates of mild, moderate, severe, and fatal reactions? Are recommendations that have been made (elimination of powder from gloves, reduction of latex in parts of devices that have body contact) having a mitigating effect? Medical device epidemiology has begun to answer these questions.⁴⁻⁸ The rest of this chapter will present some of the challenges to, and successes of, good

medical device epidemiology, emphasizing features that differ from medications epidemiology.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY RESEARCH

WHAT ARE MEDICAL DEVICES?

The term “medical devices” covers such a broad range of entities that it is difficult to define simply. However, in a regulated environment, one way of defining medical devices is to resort to a legal definition. The definition, for example, given by the United States 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.⁹

The definition offered by the European Union is somewhat different:

- (a) “medical device” means any instrument, apparatus, appliance, material or other article, whether used alone or in combination, including the software necessary for its proper application intended by the manufacturer to be used for human beings for the purpose of:
- diagnosis, prevention, monitoring, treatment or alleviation of disease,
 - diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap,

- investigation, replacement or modification of the anatomy or of a physiological process,
- control of conception,

and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means;

- (b) "accessory" means an article which whilst not being a device is intended specifically by its manufacturer to be used together with a device to enable it to be used in accordance with the use of the device intended by the manufacturer of the device;
- (c) "device used for in vitro diagnosis" means any device which is a reagent, reagent product, kit, instrument, equipment or system, whether used alone or in combination, intended by the manufacturer to be used in vitro for the examination of samples derived from the human body with a view to providing information on the physiological state, state of health or disease, or congenital abnormality thereof....¹⁰

The Canadian government's definition of a medical device is

any article, instrument, apparatus, or contrivance, including a component, part or accessory of one, that is manufactured, sold or represented for use in:

- (a) the diagnosis, treatment, mitigation, or prevention of a disease, disorder or abnormal physical state, or its symptoms, in a human being;
- (b) the restoration, correction or modification of a body function or the body structure of a human being;
- (c) the diagnosis of pregnancy in human beings, or
- (d) the care of a human being during pregnancy and at and after the birth of the child including care of the child.

It includes a contraceptive device but does not include a drug....¹¹

In Australia, the regulatory definition is

... therapeutic goods consisting of an instrument, apparatus, appliance material or other article (whether used alone or in combination), together with any accessories or software required for its

proper function, which does not achieve its principal intended action by pharmacological, chemical immunological or metabolic means though it may be assisted in its function by such means.¹¹

The Japanese government's definition is

... instruments and apparatus which are intended for use in the diagnosis, cure or prevention of diseases in man or animals, or intended to affect the structure or any function of the body of man or other animals, and which are designated by Cabinet Order.¹¹

Although the wording and level of detail is quite different for these five definitions, the overall meaning is obviously quite similar.

The main regulatory purpose of categorizing medical devices is to vary the level of premarket information required before marketing may begin, according to the potential risk posed by the particular device. This is an approach used in the European Union,¹⁰ Japan,¹² United States,¹³ Canada,¹⁴ and Australia,¹⁵ and encouraged by the Global Harmonization Task Force.¹⁶ The highest level of control is to require randomized controlled clinical trials before applying for approval to market the device. This is, obviously, 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 with an international standard recognized by the regulatory agency. 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.

Once a device is marketed, sponsors must follow Good Manufacturing Practices and monitor the safety of their products. They are subject to inspections by the regulatory agency or a proxy. The safety of most devices is monitored by keeping

a complaint file and forwarding the worst types of adverse event to the regulatory agency. More intensive efforts involve bench testing returned products, offering free replacements for returned products, and sponsoring of registries of users.

Regulatory agencies participating in the Global Harmonization Task Force monitor the safety of devices by reviewing adverse event reports from users, sponsors, or the scientific literature.¹¹ The US FDA also has a regulatory tool unique among members of the Global Harmonization Task Force;¹¹ 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.¹⁷

The nature of the medical device manufacturers themselves also varies extensively, from ownership by large pharmaceutical companies, to long term establishment as large developers of cutting-edge technology, to very small entrepreneurial operations. This variation in size naturally results in variations in safety and epidemiology expertise, as well. The large number of small companies involved is reflected organizationally at the US FDA by the Division of Small Manufacturers’ Assistance in the Center for Devices and Radiological Health, which is charged with helping small companies understand regulatory requirements.¹⁸

While governments classify medical devices for the sake of varying the level of control over different categories, epidemiologists may find other classification schemes to be more useful, e.g., the different use patterns of devices (see Table 41.1). Distinct epidemiologic problems, in the sense of hypotheses to be tested and challenges to study validity, derive from each of these categorizations. Table 41.2 shows a classification scheme that combines elements of the use patterns above, with consequent study hypotheses that might be entertained, challenges to validity, and example devices. There are several types of study hypothesis that are appropriate for epidemiologic studies of medical devices. Short term safety and efficacy issues apply to all the device categories listed in Table 41.2, except for diagnostics. Long term safety and efficacy hypotheses readily apply

Table 41.1. Spectrum of different use patterns for devices

—Number of patients exposed to a particular device	<ul style="list-style-type: none"> ● one ● multiple
—Number of times a particular device is used	<ul style="list-style-type: none"> ● once ● multiple times
—Extent of direct patient exposure to the device	<ul style="list-style-type: none"> ● none (as in laboratory analysis of a specimen) ● intact skin contact ● intact mucous membrane contact ● intact endothelium contact ● penetration of tissue ● penetration by energy or particles emitted by the device (such as ultrasound or x-rays)
—Nature of device effect	<ul style="list-style-type: none"> ● mechanical ● electronic ● chemical (e.g., diagnostic devices)
—Permanence of a particular device	<ul style="list-style-type: none"> ● used for a few minutes ● used for a few days ● used up to a year ● used several years ● “permanent” (used over 10 years)
—Setting of device use	<ul style="list-style-type: none"> ● hospital ● emergency vehicle ● clinic ● long term care facility ● private home
—Device user	<ul style="list-style-type: none"> ● doctoral care provider such as physician or dentist ● nurse ● other allied health professional ● family member ● patient

to reused and durable equipment, and long term implants. Human error issues related to proper use, and perhaps maintenance, pertain to all types of device except long term implants. 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

Table 41.2. A device categorization scheme, with implications for epidemiologic study designs

Category	Applicable types of study hypothesis	Challenges to study validity	Examples
Disposable equipment	Short term safety and efficacy; human error issues related to proper use; reuse consequences.	Device use must be inferred from procedures and knowledge of standard care, if studied retrospectively in a health care facility. Otherwise, must use interviews. Device may be part of a system that includes disposable, reusable, and durable equipment.	Barium enema cuffs, gloves
Equipment reused for the same patient and discarded within the first year of use	Short and long term safety and efficacy; human error issues related to proper use; reuse consequences.	If used in a clinic, record keeping likely to be good. At home, records may be absent. Otherwise, must use interviews. Device may be part of a system that includes disposable, reusable, and durable equipment.	Extended wear soft contact lenses
Durable equipment used for many patients and many years	Short and long term safety and efficacy; human error issues related to proper use; reuse consequences.	The particular device is uniquely identifiable only if just one is used by the facility. The type of device used must be inferred from procedure codes and knowledge of standard practice. Otherwise, must use interviews. Equipment may be refurbished or certain components may have been replaced with parts from another manufacturer. Device may be part of a system that includes disposable, reusable, and durable equipment.	Ventilators
Implants that remain in the body short term	Short term safety and efficacy; human error issues related to proper installation and maintenance.	The unique device is generally not identifiable. The type of device used must be inferred from procedure codes and knowledge of standard practice. Device may be part of a system that includes disposable, reusable, and durable equipment.	Catheters
Implants that remain in the body long term	Short and long term safety and efficacy.	Operating room notes and patient charts have detailed data; forming a cohort and following patients can be challenging, especially if the followup physician is not the implanting physician. Some registries exist.	Pacemakers
Diagnostics	Validity of test result; human error issues related to proper use.	Patient charts generally report results, not type of device. Home tests may be unrecorded. Device may be part of a system that includes disposable, reusable, and durable equipment.	Home pregnancy test kits

validity of the test result are appropriate for diagnostic devices.

INDIVIDUAL SAFETY

Because randomized clinical trials have the power to demonstrate clinically significant effi-

cacy (a positive effect must be demonstrated in a substantial proportion of patients), but little power to assure safety (because adverse effects affect a much smaller fraction of patients), safety has become largely the domain of epidemiology. The body of safety data includes private studies by the device manufacturer, adverse event

reports (redacted portions are publicly available), and epidemiologic studies reported in the scientific literature. Only high risk and some other new devices are subject to a clinical trial requirement. Devices that are similar to well established devices only have to be demonstrated as meeting a standard, or being similar to a marketed device; if the predicate device was never subjected to a randomized clinical trial (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.

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 various options, such as doing nothing, taking an action, or perhaps taking an alternative action. Available epidemiologic evidence of both the effectiveness and extent of device use, as well as the availability of alternative therapies, have a bearing on the decision. For example, gloves are made of either latex or vinyl; vinyl is less likely to cause allergic reactions, but latex is far more flexible and durable and is still used in most situations.¹⁹

DIFFERENTIAL EFFECTS

Another problem that is amenable to epidemiology is the study of the differential effects of devices by some cofactor, such as gender or a concomitant therapy. In other words, epidemiology can help identify patients at higher risk of complications from a device. Two examples are the higher risk of ventilator-associated pneumonia in men and in patients receiving a paralyzing agent.²⁰

METHODOLOGIC PROBLEMS TO BE SOLVED BY PHARMACOEPIDEMOLOGY RESEARCH

RECOGNITION AND CHARACTERIZATION OF NEW SYNDROMES OR ADVERSE EVENTS

Sometimes, as with the latex barium enema cuff example, 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. An example is First Use Syndrome occurring during hemodialysis. As the name implies, this syndrome only occurs when a device, such as a dialysis membrane (which may be routinely reused many times), is brand new.^{21,22} Where controversy exists over the establishment of a definition of a new syndrome, as is the case with the possibility of a new connective tissue disease that may be associated with silicone gel breast implants, the situation is even murkier.²³ In all of these circumstances, epidemiologists can contribute to improved methodology to detect and characterize adverse events and syndromes.

INDIVIDUAL EXPOSURE ASSESSMENT

For epidemiologists, it is a given that a determination of exposure and outcome status at the level of the individual study subject is critical to making confident assessments of the relationship between the two. The general ways to do this for medical device studies are to consult a medical record or to ask the subject.

Almost all of the challenges to medical device study validity that are listed in Table 41.2 relate to individual exposure assessment. The brand, model, or 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. For instance, a barium enema cuff is likely to have been used for a barium enema procedure, but the brand and model will not have been recorded.

In the case of durable equipment, if only one was available at the facility, it is possible to obtain 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 a different manufacturer during repair or refurbishment.

For long term implants, operating room notes and patient data generally have detailed data on the implant. Registries have been formed for a variety of implants, although patients may resist registration if the implant is of a socially sensitive nature. Long term followup, 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.

Prospective studies to confirm the methodology of inferring the particulars of device use need to be conducted, especially in light of the fact that there is no national or international nomenclature for devices that is comparable to the National Drug Code for drugs. The current schemes, FDA's Medical Device Product Code and ECRI's Universal Medical Device Nomenclature SystemTM,²⁴ are based on intended use and do not specify brand, model, or other details that may be relevant, such as size or material.

When asking the 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 erectile dysfunction devices, are associated with social discomfort, so may be underreported.²⁵ The interviewer must take special care to encourage full disclosure.

As an example of some of these methodologic problems, consider the case of extended use of soft contact lenses. 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 status difficult to measure from records alone. In their studies of corneal ulcer, exposure was assessed by Poggio *et al.*²⁶ with a population survey, and by Schein *et al.*²⁷ with a structured interview of all cases and controls. The different methods used in each study were probably valid because they resulted in similar effect estimates: relative risks of 5.2 (95% confidence interval of 3.5 to 7.7) by Poggio *et al.* and 3.9 (95% confidence interval of 2.4 to 6.5) by Schein *et al.*

NATIONAL POPULATION EXPOSURE ASSESSMENT

Once a regulator has discovered a likely relationship between exposure and outcome, he or she needs to determine the extent of population exposure. Public sources of device exposure data include market, medical care claims, medical records, and population survey data. Market data firms derive their information in various ways, from polling health care providers^{28,29} to collecting device purchase information from a nationally representative set of health care facilities and providers.³⁰ Market data may express incident or prevalent exposure data. A nationally representative hospital claims database provides national incidence estimates for devices that can be adequately measured by hospital procedure codes.^{31,32} Other medical care claims records can also be a source of procedure codes that may be used to infer device use. Exposure estimates can also be made by consulting a sample of medical records or by surveying the population. The

limitations of measuring device exposure with medical data sources that are discussed above in the section on individual exposure assessment also apply to these methodologies. The Medical Device Implant Supplement to the 1988 National Health Interview Survey of households provided the first population prevalence assessment of a variety of implants.^{25,33,34} In general, these information sources are much less abundant, reliable, or detailed than comparable types of drug exposure information.

CURRENTLY AVAILABLE SOLUTIONS

This section describes observational and descriptive epidemiologic tools currently available for studying medical device use.

SAFETY SURVEILLANCE

In the US, general safety surveillance is done by FDA through voluntary and mandatory reporting of suspected adverse events due to devices. Manufacturers must report malfunctions, serious injuries, and deaths. Health care facilities must report deaths.³⁵ Anyone can make a voluntary report. The FDA receives reports of about 100 000 adverse device events per year³⁶ (compared to about 160 000 for drugs).³⁷ They are recorded and reviewed in a manner similar to drug reports, although the reviewers are largely nurses, as opposed to pharmacists, because nurses generally have the most extensive clinical experience with devices. Reports are assigned to individual reviewers based largely on the organ system related to the device's intended use, so that each reviewer is responsible for a large variety of types of device. Reviewers may ask the reporters for more information. FDA has taken actions that include recalls, public alerts, required changes to device labeling, and required epidemiologic studies.

Although medical device adverse event reporting is also required by Canada, Japan, Australia, and the European Union, very little public information is available about these systems, besides what is specified in their regulations. All of these countries have mandatory reporting by

device sponsors and voluntary reporting by anyone. The types of reportable event are very similar to those in the US. None of these other countries require reporting by health care facilities.¹¹

The general discussions of this type of system found elsewhere in this book 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 surveys conducted by the US National Center for Health Statistics have provided a variety of data on medical devices, including the Medical Device Implant Supplement to the 1988 National Health Interview Survey³⁸ mentioned above, the 1988 National Maternal and Infant Health Survey,³⁹ the National Health and Nutrition Examination Survey series,⁴⁰ the 1993 Mortality Followback Survey,⁴¹ and the 1996 National Home and Hospice Care Survey.⁴² These are summarized in Table 41.3.

The Medical Device Implant Supplement to the 1988 National Health Interview Survey³⁸ mentioned above was used to generate the first overall national prevalence estimates of implants.³³ More extensive reports were generated for the particular devices. Examples include one on breast implants,²⁵ which demonstrated that using a general medically oriented screening question asked of a household representative may not elicit all socially sensitive or cosmetic implants for all household members. Another example was an analysis of pacemakers,³⁴ showing that there were about 456 000 noninstitutionalized implant recipients; about 15% of the prevalent implants were replacements. Prevalence rose steeply with recipient age and was higher for men, whites, and those reporting an activity limitation. Geographic region, income, and educational level were not related to pacemaker prevalence.

The 1988 National Maternal and Infant Health Survey was used to generate national statistics on home pregnancy test use among women who carried their pregnancies to term.⁴³ It was estimated that a third of such women used the test,

Table 41.3. Some surveys conducted by the US National Center for Health Statistics that include medical device information

Title	Device information	Sample size
Medical Device Implant Supplement to the 1988 National Health Interview Survey	“The National Health Interview Survey is a [periodic] cross-sectional household interview survey. Sampling and interviewing are continuous throughout each year. The sampling plan follows a multistage area probability design that permits the representative sampling of households.” ³⁸ 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. ³³	The sample included 122 000 members of about 47 000 households. The net MDIS response rate was 92%. ³³
1988 National Maternal and Infant Health Survey. ³⁹	“The NMIHS is based on questionnaires administered to nationally representative samples of mothers with live births, stillbirths, and infant deaths during 1988 and to physicians, hospitals, and other medical care providers associated with those outcomes.” Data from the mothers and care providers include prenatal and perinatal tests and procedures, such as ultrasound examination. “A 1991 longitudinal followup to the NMIHS was conducted to obtain additional information about respondents from the 1988 survey.” Information about some devices, such as home apnea monitors, is included.	“The survey is based on 10 000 live births, 4000 fetal deaths, and 6000 infant deaths.”
National Health and Nutrition Examination Survey series ⁴⁰	“The survey consists of standardized physical examinations, health interviews, and laboratory testing to provide data on the health and nutritional status of the population. The survey produces data on a wide range of diseases and conditions, risk factors and health behaviors, physiological and body measurements, and environmental exposures ... The latex allergy test was measured on a nationally representative sample of persons aged 17–60 years during the 1988–91 (phase 1) portion of the NHANES III survey. Although NHANES survey data does not include specific information on symptoms due to latex, participants will be classified as sensitized or not sensitized based on the level of detectable latex-specific IgE in sera, using cutoffs designated by the manufacturers.”	30 000 adults and children participated in the most recent survey, NHANES III.
1993 Mortality Followback Survey ⁴¹	The subjects were a random sample of deaths that occurred in 1993 in the US (except South Dakota), stratified by age, race, and cause of death. This survey was “the first study since the 1980’s to examine detailed patterns of mortality by supplementing the information provided through death certificates with interviews of next of kin. The survey allows for the examination of trends in mortality, differences by income and education, risk factors and causes of death, and health care utilization in the last year of life ...” “[O]ver 1100 data items [were] collected on such topics as: underlying and multiple cause of death; utilization of health care services; health care payment sources; access to care; health conditions; ... use of assistive devices; medical implants; ... sociodemographic information; place of death and circumstances of injury; ... life events and activities; problem behaviors; income and assets; and respondent background characteristics.” It provides information on devices used in the home (blood glucose meters, dialysis machines, etc.).	Almost 23 000 subjects.
1996 National Home and Hospice Care Survey ⁴²	“The sampling frame for the 1996 National Home and Hospice Care Survey (NHHCS) consisted of 16 700 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 1091 “agencies providing home health or hospice care services at the time of the survey ... Of [these], 1053 (97 percent) agreed to participate ...”

and that they were more likely to be white, married, older, more educated, and wealthier than the other mothers.

Hefflin used the 1993 Mortality Followback Survey to conclude that “older persons who had initial pacemakers implanted during their final year of life ... [were] relatively independent, physically functional candidates who frequently died

unexpectedly,” indicating that expert guidelines for pacemaker implantation were generally followed. Eighteen percent of the estimated 79 000 elderly recipients of pacemakers who had died in 1993 had received the implant for the first time during their final year of life.⁴⁴

Ad hoc surveys are also used to study specific devices, such as latex gloves. Table 41.4 shows the

Table 41.4. Features of five example studies of sensitivity to latex gloves

Author and year of study	Surveyed population	Latex sensitivity assessment method	Sensitivity rate	Sensitivity predictors
Turjanmaa 1987 ⁵	All operating room and laboratory employees in a Finnish hospital. Also, 130 consecutive patients who were due for routine scratch testing.	Scratch test for latex, vinyl, 20 common inhalant allergens with histamine, and diluent. Glove use test with latex and with vinyl.	Employees: 2.8%. Patients: 0.8%. $p < 0.01$	Atopy, hand eczema, operating room work, physician.
Sussman <i>et al.</i> 1995 ⁴	All 71 housekeeping staff in a Canadian hospital. 70% participated. All wore unpowdered, unflocked latex gloves for 25 to 30 hours per week.	Questionnaire screen; positives were skin tested with latex, eight common inhalant and food allergens, and histamine.	8%.	
Kaczmarek <i>et al.</i> 1996 ⁶	915 emergency room workers (nurses, physicians, emergency medical technicians) in nine hospitals. 42% completed all sensitivity measures.	Questionnaire and test of serum for IgE antibodies to latex.	6%, by serum test.	Any allergy history, nonwhite race.
Brown <i>et al.</i> 1998 ⁷	All 171 health care employees of an anesthesiology and critical care department of a hospital. Excluded provocation tests for those with recent unstable asthma or interfering medications, or pregnancy. 168 provided history and blood; 154 provided the skin test.	History; blood test for latex, three fruits, and a combination of eight common allergens; skin test to latex, glycerinated saline, histamine, and nonammoniated latex; and glove provocation test.	20% by history. 12.5% by blood test.	Atopy; food allergies; specific allergies to banana, kiwi, or avocado.
Sussman <i>et al.</i> 1998 ⁸	Phase 1 : 2062 health care workers at two sites of a hospital. 1351 participated. Phase 2: introduction of powder-free gloves at one site; continuance of powdered gloves at the other. Of 479 eligible workers negative at baseline, 435 completed followup.	1 year prospective followup. Skin test with latex, noncompounded ammoniated latex, several food and inhalant allergens, and saline used to measure new sensitivity.	1.0% in powdered; 0.9% in nonpowdered glove group.	Atopy.

features of five example studies.⁴⁻⁸ All of these latex studies were occupationally based, on the assumption that sensitivity rates are higher in these groups. Studies directed to understanding risk factors and testing candidate interventions are statistically relatively powerful in these occupational groups, and the results presumably apply to patients.

As described in an earlier section of this chapter, *ad hoc* surveys were required to measure the use and complications from extended wear soft contact lenses.²⁶ Poggio *et al.* 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.

REGISTRIES

Registries are sometimes formed to establish cohorts of patients with particular device exposures. Several registries form the basis of pacemaker studies. Examples include the Implantable Lead Registry, formed by six North American hospitals in 1979,⁴⁵ the Fyn County registry in Denmark formed in 1964,⁴⁶ and the Danish Pacemaker Register for all of Denmark begun in 1982.⁴⁷ In addition, Kawanishi *et al.*⁴⁸ used data from four registries: the Bilitch Registry of pacemaker pulse generators, operating from 1973 to 1993, with three to six sites, 22 786 devices, and 16 903 patients; the Implantable Lead Registry of pacemaker leads, operating from 1979 to 1989, with the same sites as the Bilitch Registry and 7311 patients; the United States Veterans Administration Registry of Pacemaker Leads, still operating in 1992 at 182 facilities, with 8612 patients; and the Cleveland Clinic Lead Registry, operating from 1980 to 1991. Device life can be estimated by type, brand, model, and reason for replacement.^{45,47} If the registry is population based, the incidence of device implantation, prevalence of implant recipients, and relative mortality rates for recipients can also be calculated.⁴⁶

MEDICAL CARE CLAIMS

An example of a large medical care claims database suitable for medical device epidemiology

is the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS), which has been made available by the Agency for Healthcare Research and Quality (formerly the Agency for Health Care Policy and Research). This database is updated periodically.³¹

The HCUP Nationwide Inpatient Sample (NIS) contains all-payer data on hospital inpatient stays from selected States. Each year of the NIS provides information on approximately 5 million to 7.1 million inpatient stays from over 900 hospitals. All discharges from sampled hospitals are included in the NIS database.

The NIS contains patient-level clinical and resource use information included in a typical discharge abstract. The NIS can be linked directly to hospital-level data from the American Hospital Association (AHA) Annual Survey of Hospitals and to county-level data from the Health Resources and Services Administration Bureau of Health Professions' Area Resource File (ARF), except for hospitals from Georgia, Hawaii, Kansas, South Carolina, and Tennessee.

The NIS is designed to approximate a 20-percent sample of U.S. community hospitals.^{31,49}

NIS data are available for 1988 through 1996. Since the data consist of discharge claims, studies of devices depend on the adequacy of procedure codes for capturing device exposure. Within this limitation, studies of the extent of use, adverse events in the same hospitalization, or longitudinal trends can be performed, as was recently done for pacemaker implantations.³² US census data adjusted to the year 1992 was used along with the national pacemaker implantation estimates from NIS, to calculate 5 year age group and gender specific implantation rates for the US. The investigators were also able to analyze the principal diagnoses, associated diagnoses, whether the implant was a replacement, and death before discharge.

MEDICAL RECORDS

Medical records are a resource for medical device epidemiology to the extent that they note the use of devices. Medical record systems vary from entirely paper based, to partially systematized, to fully automated.

Paper based medical record systems were used to measure outcomes for some pacemaker studies. Rubin *et al.*⁵⁰ reported their retrospective review of 287 patients at one clinic in the southeastern US who had been actively followed. They calculated generator life, complications, replacements, and patient mortality. Another example is from Mueller *et al.*,⁵¹ who reported on patients who had been implanted and regularly followed at an Austrian cardiology department. They were able to analyze by gender, age at implantation, indication, patient survival (compared to the general Austrian population), and cause of death.

A partially systematized structure is used at the Mayo Clinic, where another pacemaker survival study was conducted.⁵² The Mayo Clinic has access to all data regarding medical care provided in Olmsted County, MN, USA.⁵³ A card system was available to identify pacemaker recipients, but other medical information was obtainable from the full paper medical record. Shen *et al.* defined a group of patients and followed them through time with the medical records and death certificates. Patient survival was calculated in comparison to the North Central white population at a comparable time, and stratified by demographics, clinical history, heart disease type, and some comorbidities. Another Mayo Clinic study with a similar data collection strategy investigated transtracheal oxygen catheters.⁵⁴ Records of patients receiving the device were reviewed for demographics, reason for placement, complications, reason for removal, and duration of use.

Harvard Pilgrim Health Care is a managed care provider with a fully automated patient record system (see Chapter 17).

As demonstrated by the examples, 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. A case in point is soft contact lens use, as described earlier in this chapter,^{26,27} where only the subject can reliably provide the exposure (length of wear times) and some covariate information (such as cleaning practices).

COMBINATIONS OF TECHNIQUES

At times, as for other epidemiologic situations, a combination of study techniques may be used to strengthen the design. In a study of pacemaker use, Greenspan *et al.*⁵⁵ presumably used Medicare claims obtained by a Professional Standards Review Organization in the US to identify patients; the remainder of the information was collected via extensive chart review. The authors evaluated the appropriateness of implantation of Medicare reimbursable pacemakers.

Another example of multiple techniques is the use of different survey types as well as complementary study designs (case-control and cohort) to address the same primary and complementary secondary questions. 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 complementary epidemiologic studies were designed together and incorporated both a random sample survey of a population and of all ophthalmologists serving the population,²⁶ as well as case selection with both hospital and population controls.²⁷ The first study established the incidence rate of corneal ulcer and the second demonstrated several risk factors. It appeared that risk rises with each day of wear but varies among individuals (some people cannot tolerate even one night of extended wear, while others experience no trouble with a week). Based on these studies, FDA found it prudent to have eye care professionals set individual patient maximum wear times, subject to an overall limit of 7 days.⁵⁶

THE FUTURE

TRENDS IN MEDICAL DEVICE TECHNOLOGY

Medical devices are becoming more sophisticated, smaller, more precise, and easier to use in the home.⁵⁷ As applications increase, current medical device nomenclatures that are based on intended use are becoming cumbersome. An international effort is currently under way to standardize and update the nomenclature.^{24,58} A useful future step would be to use this nomenclature in individual patient records to help define device exposure.

Another trend is the ever speedier introduction of new generations of devices.⁵⁹ In some cases, new technology is in use before an epidemiologic study of the older technology can be completed. This problem will need to be addressed by faster techniques, such as quicker assemblage of an exposed cohort after a device is introduced to the market, for observing long term safety and effectiveness.

TRENDS IN EPIDEMIOLOGIC RESOURCES

Increasing uniformity and automation of medical records will improve the prospects for medical device epidemiology.⁶⁰⁻⁶² Some observers believe the Internet will enhance the conduct of records based studies.⁶³⁻⁶⁶ However, proposed privacy laws could greatly restrict access by researchers.^{61,67} Acceptable standard methods may be developed to address confidentiality concerns while allowing rigorous epidemiologic research.⁶⁸

TRENDS IN MEDICAL DEVICE EPIDEMIOLOGY

Awareness of medical devices as an object of epidemiologic study has been growing. It is my hope that as the field grows, recognition of medical devices as an influence on health will expand, allowing major advances in methodology.

ACKNOWLEDGEMENTS

The author is grateful to Roscoe M. Moore, Jr., DVM, MPH, PhD, DSc, for guidance and inspiration, and thanks Mr. Greg Yoder for editorial assistance with an early draft.

REFERENCES

1. Feczko PJ, Simms SM, Bakirci N. Fatal hypersensitivity reaction during a barium enema. *Am J Roentgenol* 1989; **153**: 275-6.
2. Feczko PJ. Increased frequency of reactions to contrast materials during gastrointestinal studies. *Radiology* 1990; **174**: 367-8.
3. Tomazic VJ, Withrow TJ, Fisher BR, Dillard SF. Short analytical review: latex-associated allergies and anaphylactic reactions. *Clin Immunol Immunopathol* 1992; **64**: 89-97.
4. Sussman GL, Lem D, Liss G, Beezhold D. Latex allergy in housekeeping personnel. *Ann Allergy Asthma Immunol* 1995; **74**: 415-8.
5. Turjanmaa K. Incidence of immediate allergy to latex gloves in hospital personnel. *Contact Dermatitis* 1987; **17**: 270-5.
6. Kaczmarek RG, Silverman BG, Gross TP, Hamilton RG, Kessler E, Arrowsmith-Lowe JT, Moore RM. Prevalence of latex-specific IgE antibodies in hospital personnel. *Ann Allergy Asthma Immunol* 1996; **76**: 51-6.
7. Brown RH, Schauble JF, Hamilton RG. Prevalence of latex allergy among anesthesiologists: identification of sensitized but asymptomatic individuals. *Anesthesiology* 1998; **89**: 292-9.
8. Sussman GL, Liss GM, Deal K, Brown S, Cividino M, Siu S, Beezhold DH, Smith G, Swanson MC, Yunginger J, Douglas A, Holness DL, Lebert P, Keith P, Wasserman S, Turjanmaa K. Incidence of latex sensitization among latex glove users. *J Allergy Clin Immunol* 1998; **101**: 171-8.
9. United States Congress (1998) Federal Food, Drug, and Cosmetic Act, Sec. 201 [321]. Definitions (h).
10. The Council of the European Communities Directive 93/42/EEC of 14 June 1993 concerning medical devices. Document No. 393L0042. http://europa.eu.int/eur-lex/en/lif/dat/1993/en_393L0042.html.
11. Comparison of the Device Adverse Report Systems in USA, Europe, Canada, Australia, and Japan. Global Harmonization Task Force Study Group 2—Medical Device Vigilance—Post-Market Surveillance N6: October 1997 Working Draft. <http://www.ghtf.org/general/inventorysg2/regover.html>.
12. Ensuring the Safety of Drugs and Medical Devices. White Paper: Annual Report on Health and

- Welfare, Book 1, Part 2, Chapter 3, Section 4. http://www.mhw.go.jp/english/white_p/index.html.
13. United States Code of Federal Regulations, Title 21—Food and Drugs, Part 860—Medical Device Classification Procedures. 1 April 1998.
 14. Therapeutic Products Programme, Health Canada. Frequently Asked Questions and Answers Concerning the New Medical Devices Regulations. 1 October, 1998. <http://www.hc-sc.gc.ca/hpb-dgms/therapeut/htmleng/guidmd.html>.
 15. Therapeutic Goods Administration, Australia. Australian Medical Device Requirements under the Therapeutic Goods Act 1989 (version 4) “DR4”. <http://www.health.gov.au/tga/dr4.htm>.
 16. Global Harmonization Task Force Study Group 1. Recommendation on Medical Devices Classification. Draft of 15 January, 1999. <http://www.ghtf.org/general/inventoriesgl/n015r4-sgl.pdf>.
 17. Food and Drug Administration Modernization Act of 1997. Sec. 212. Postmarket Surveillance. <http://www.fda.gov/cder/guidance/s830enr.txt>.
 18. Center for Devices and Radiological Health. Division of Small Manufacturers Assistance (DSMA) (1999). <http://www.fda.gov/cdrh/dsma/dsmamain.html>.
 19. Therapeutic Products Programme, Health Canada. Glove barrier protection and latex allergy: minimizing the risk. *Med Devices Bull* 1998; **4**: 16–7.
 20. Cook DJ, Walter SD, Cook RJ, Griffith LE, Guyatt GH, Leasa D, Jaeschke RZ, Brun-Buisson C. Incidence of and risk factors for ventilator-associated pneumonia in critically ill patients. *Ann Intern Med* 1998; **129**: 433–40.
 21. Nicholls AJ, Platts MM. Anaphylactoid reactions due to haemodialysis, haemofiltration, or membrane plasma separation. *Br Med J* 1982; **285**: 1607–9.
 22. Hakim RM, Breillatt J, Lazarus JM, Port FK. Complement activation and hypersensitivity reactions to dialysis membranes. *New Engl J Med* 1984; **311**: 878–82.
 23. 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.
 24. Testimony by ECRI to the National Committee on Vital and Health Statistics, Subcommittee on Health Data Needs, Standards and Security, 11 April 1997. <http://aspe.os.dhhs.gov/ncvhs/97041620.htm>.
 25. Bright RA, Jeng LL, Moore RM. National survey of self-reported breast implants: 1988 estimates. *J Long-Term Effects Med Implants* 1993; **3**: 81–9.
 26. Poggio EC, Glynn RJ, Schein OD, Seddon JM, Shannon MJ, Scardino VA, Kenyon KR. The incidence of ulcerative keratitis among users of daily-wear and extended-wear soft contact lenses. *New Engl J Med* 1989; **321**, 779–83.
 27. 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. Microbial Keratitis Study Group. *New Engl J Med* 1989; **321**: 773–8.
 28. Bernstein AD, Parsonnet V. Survey of cardiac pacing and defibrillation in the United States in 1993. *Am J Cardiol* 1996; **78**: 187–96.
 29. Parsonnet V, Bernstein AD, Galasso D. Cardiac pacing practices in the United States in 1985. *Am J Cardiol* 1998; **62**: 71–7.
 30. IMS America. *Hospital Supply Index* 1998. <http://www.imshealth.com>.
 31. AHRQ. *Healthcare Cost and Utilization Project (HCUP), 1988–97: A Federal–State–Industry Partnership in Health Data*. Rockville, MD, Agency for Healthcare Research and Quality, 2000. <http://www.ahrq.gov/data/hcup-pkt.htm>.
 32. Daley WR, Kaczmarek RG. The epidemiology of cardiac pacemakers in the older US population. *J Am Geriatr Soc* 1998; **46**: 1016–9.
 33. Moss AJ, Hamburger S, Moore RM, Jeng LL, Howie LJ. *Use of Selected Medical Device Implants in the United States, 1988. Advance Data from Vital and Health Statistics; No. 191*. Hyattsville, MD: National Center for Health Statistics.
 34. 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.
 35. United States Federal Government. *Title 21 Code of Federal Regulations, Chapter I, Part 803—Medical Device Reporting*. 1998.
 36. White GG, Weick-Brady MD. Improving patient care by reporting problems with medical devices. A MedWatch Continuing Education Article. 1997. <http://www.fda.gov/medwatch/articles.htm>.
 37. Annual adverse drug experience report: 1996. 30 October 1997. <http://www.fda.gov/cder/dpe/annrep96/index.htm>.
 38. National Health Interview Survey (NHIS). <http://www.cdc.gov/nchswww/about/major/nhis/nhis.htm>.
 39. National Maternal and Infant Health Survey. <http://www.cdc.gov/nchswww/about/major/nmihs/abnimhs.htm>.
 40. National Health and Nutrition Examination Survey series. <http://www.cdc.gov/nchswww/about/major/nhanes/nhanes.htm>.
 41. National Mortality Followback Survey (NMFS). <http://www.cdc.gov/nchswww/about/major/nmfs/nmfs.htm>.
 42. National Home and Hospice Care Survey (NHHCS). <http://www.cdc.gov/nchswww/about/major/nhhcsd/nhhcsd.htm>.
 43. 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**: 11–3.
44. Hefflin BJ. Final-year-of life pacemaker recipients. *J Am Geriatr Soc* 1998; **46**: 1396–1400.
45. Furman S, Benedek ZM. Survival of implantable pacemaker leads. The Implantable Lead Registry. *Pacing Clin Electrophysiol* 1990; **13**: 1910–4.
46. Andersen C, Green A, Madsen GM, Arnsbo P. The epidemiology of pacemaker implantations in Fyn County, Denmark. *Pacing Clin Electrophysiol* 1991; **14**: 1614–21.
47. Moller M, Arnsbo P. Appraisal of pacing lead performance from the Danish Pacemaker Register. *Pacing Clin Electrophysiol* 1996; **19**: 1327–36.
48. Kawanishi DT, Song S, Furman S, Parsonnet V, Pioger G, Petitot JC, Godin JF. Failure rates of leads, pulse generators, and programmers have not diminished over the last 20 years: formal monitoring of performance is still needed. BILITCH Registry and STIMAREC. *Pacing Clin Electrophysiol* 1996; **19**: 1819–23.
49. Introduction to the HCUP Nationwide Inpatient Sample (NIS). Rockville, MD, Agency for Healthcare Research and Quality, 2000. <http://www.ahrq.gov/data/hcup/nisintro.htm>.
50. Rubin JW, Killam HAW, Moore HV, Ellison RG. Permanent cardiac pacemakers: twelve-year experience with 287 patients. *Ann Thorac Surg* 1976; **22**: 74–9.
51. Mueller C, Cernin J, Glogar D, Laczkovics A, Mayr H, Scheibelhofer W, Schmidinger H, Schuster E, Sedlacek K, Kaliman J. Survival rate and causes of death in patients with pacemakers: dependence on symptoms leading to pacemaker implantation. *Eur Heart J* 1988; **9**: 1003–9.
52. Shen WK, Hammill SC, Hayes DL, Packer DL, Bailey KR, Ballard DJ, Gersh BJ. Long-term survival after pacemaker implantation for heart block in patients ≥ 65 years. *Am J Cardiol* 1994; **74**: 560–4.
53. Kurland LT, Molgaard CA. The patient record in epidemiology. *Sci Am* 1981; **245**: 54–63.
54. Orvidas LJ, Kasperbauer JL, Staats BA, Olsen KD. Long-term clinical experience with transtracheal oxygen catheters. *Mayo Clin Proc* 1998; **73**: 739–44.
55. Greenspan AM, Kay HR, Berger BC, Greenberg RM, Greenspon AJ, Gaughan MJS. Incidence of unwarranted implantation of permanent cardiac pacemakers in a large medical population. *New Engl J Med* 1988; **318**: 158–63.
56. Villforth JC. *New FDA Recommendations and Results of Contact Lens Study*, FDA letter to eyecare practitioners. 30 May 1989.
57. Spera G. CDRH predicts tomorrow's top technologies. *Med Device Diagnost Industry* 1998; **20**: 18, 20–1.
58. ECRI. ECRI to Work with the U.S. Food and Drug Administration Center for Devices and Radiological Health to Harmonize Medical Device Nomenclature Systems. 1998. <http://www.healthcare.ecri.org>.
59. Padley K. Guidant's defibrillator wins race to market. *Saint Paul Pioneer Press*, 10 September 1998: 1.
60. Harman J. Topics for our times: New health care data—new horizons for public health. *Am J Public Health* 1998; **88**: 1019–21.
61. Coyne MM Jr. Information versus privacy: an imaginative look at the future. *Med Device Diagnost Industry* 1998; **20**: 24–6.
62. Longe K. How to implement a UPN system. *Med Device Diagnost Industry* 1998; **20**: 108–12.
63. Hjelm NM, Tong FFK. Patients' records on the Internet: a boost for evidence-based medicine. *Lancet* 1998; **351**: 1751–2.
64. Goldberg HI, Tarczy-Hornoch P, Stephens K, Larson EB, LoGerfo JP. Internet access to patients' records. *Lancet* 1998; **351**: 1811.
65. de Groen PC, Barry JA, Schaller WJ. Applying World Wide Web technology to the study of patients with rare diseases. *Ann Intern Med* 1998; **129**: 107–13.
66. Shortliffe EH. Health care and the next generation Internet. *Ann Intern Med* 1998; **129**: 138–40.
67. Melton LJ III. The threat to medical-records research. *New Engl J Med* 1997; **337**: 1466–70.
68. Rabinowitz J. A method for preserving confidentiality when linking computerized registries. *Am J Public Health* 1998; **88**: 836.

42

Special Considerations in Studies of Drug-induced Birth Defects

ALLEN A. MITCHELL

Slope Epidemiology Unit, Departments of Epidemiology and Pediatrics, Boston University Schools of Public Health and Medicine, Brookline, MA, USA

INTRODUCTION

Teratogenesis is a very different phenomenon from other drug-induced hazards, and it therefore requires special consideration. Although the fetus may experience a wide range of adverse effects as a result of antenatal drug exposure, such as mental/motor deficits and learning and behavioral problems, this chapter will confine itself to issues surrounding drug-induced physical malformations.

THE NATURE OF BIRTH DEFECTS AND THEIR RELATION TO DRUGS

Birth defects are part of the human condition, having been observed throughout history. Major birth defects, typically defined as those that are life threatening, require major surgery, or present a significant disability, affect approximately 3–4% of liveborn infants.^{1,2} Minor malformations are of lesser clinical importance, and estimates of their prevalence vary considerably because of substantial differences in definition and detection.

Over the centuries, the “deformities” that characterize most birth defects have been viewed as a punishment to the mother or family for some fault on their part. This view was undoubtedly reinforced by the rarity of birth defects, their unpredictable occurrence, and the absence of known causes. Perhaps because these factors have not changed very much over time, elements of this primitive view persist today, largely in the form of guilt. Parents tend to search their memories to identify some factor—any factor—that might account for their misfortune. In developed societies, attention often focuses on drugs taken in pregnancy.

This concern, of course, is not without foundation. Less than 60 years ago, it was believed that the placenta protected the fetus from noxious agents. That belief was shattered in 1941 by the recognition that maternal rubella infection in pregnancy produced a distinctive pattern of birth defects among exposed infants.³ Two decades later, the thalidomide disaster demonstrated that drugs, too, could be teratogenic.⁴ Many thousands

of infants were born with major limb reductions and other defects, and the tragedy of this epidemic was etched into the consciousness of medical practitioners and the public alike. In the years that followed, other drugs were shown to be teratogenic, ranging from phenytoin to isotretinoin. Many additional drugs were alleged to be teratogenic, and although most of those allegations were unsupported by subsequent studies, they served to reinforce the general concern about the teratogenic effects of marketed drugs.

Teratogenesis is a unique kind of adverse drug effect, since it affects an organism (the fetus) other than the one for whom the drug was intended (the mother). In a benefit/risk consideration, the fetus may at best indirectly benefit from a medication given to its mother (e.g., by an improvement in the mother's health), but the fetus alone is at risk for birth defects. That "innocent bystander" status of the fetus raises profound medical, moral, and legal issues (see also Chapters 5 and 9). It also poses serious concerns about the consequences of allegations that a given drug may be teratogenic.

CAN TERATOGENESIS BE PREDICTED PRIOR TO MARKETING A DRUG?

Under ideal circumstances, one would identify the teratogenic potential of drugs before they were used in humans. Unfortunately, our ignorance about the basic mechanisms of organ formation has constrained development of predictive *in vitro* tests. Testing in animals may be helpful in certain circumstances; for example, vitamin A and its congeners produce consistent patterns of malformations across many different species. However, for most known human teratogens (including thalidomide), the results of animal tests vary so much as to seriously limit their predictive value.⁵ Further, understanding of the structure/activity relationships of a particular agent can help predict that drug's efficacy and adverse reactions, but it does not necessarily help predict its teratogenic potential. Because there are no theoretical, *in vitro*, or animal models that can reliably provide meaningful information about the likelihood of fetal risk, we are usually completely unaware of a given drug's teratogenic potential when it is first used in humans.

Clinical premarketing studies cannot be expected to provide this information either. Until recently, women of childbearing age were excluded from early clinical studies, specifically because of concerns about potential teratogenicity. Newer guidelines are designed to reverse these exclusions, but women chosen to be enrolled in clinical studies will, appropriately, be those at minimal risk of becoming pregnant. Information derived from experience among nonpregnant adult women is not informative when the concern is teratogenesis, and might even provide a false sense of reassurance—recall that thalidomide was used as a sedative in pregnant women specifically because of its "safety profile" in nonpregnant adults.

A Note about Nonprescription Drugs

Nonprescription (or over-the-counter, OTC) drugs present a unique situation. Whatever caution physicians might exercise in their prescribing of drugs to pregnant women, they have little control over what consumers purchase over the counter or obtain from their friends, relatives, and neighbors. It can reasonably be assumed that women, like their physicians, perceive nonprescription drugs as being safer than prescription products, and therefore may be more willing to use OTC agents in pregnancy. This perception is based on the fact that a switch from prescription to OTC status implies a history of wide use and safety. However, because there is little systematic information available on the human teratogenicity of most prescription drugs, the process of switching to OTC availability rarely takes account of a drug's risk or safety with respect to the fetus. This is particularly the case for drugs that have been available OTC for many years. As noted below, teratogenicity is more difficult to assess when a drug is used without prescription, and it is ironic that we may know less about the teratogenic hazard of drugs available OTC than we do about drugs available only by prescription.

Thus, when prescription drugs are first licensed for general use, and for most OTC drugs, we lack an understanding of their teratogenic effects and we typically lack an understanding even of their teratogenic potential; the opportunity to gain

such knowledge comes (if it comes at all) from postmarketing experience, where pharmacoepidemiology can and must play a crucial role.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY RESEARCH

Like other adverse drug effects, teratogenesis is a critical aspect of a drug's benefit/risk profile, and such information obviously should be available to prescribers and consumers. Unlike other adverse drug effects, however, teratogenesis raises uniquely important and controversial clinical issues. First, the fetus is the "innocent bystander" with respect to its mother's therapy. Second, teratogenesis is not a concern limited to women who are pregnant when drug treatment is initiated; it must also be a concern among women who might become pregnant after a drug is prescribed. 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

Broadly, there are three approaches applied to the few drugs known to be teratogenic. In rare instances, such as was the case for thalidomide in most countries, the drug was prohibited from the general market; this approach was justified by the large absolute risk to the fetus, and is feasible in instances where the drug does not offer important or unique therapeutic benefits.

For most known teratogens, such as phenytoin and valproic acid, the absolute risk to the fetus is modestly elevated and the drug is felt to fill an important clinical need. For such drugs, information about teratogenicity is provided to physicians, who are expected to discuss the benefits and risks with their patients, who then can make informed decisions about their drug treatment. In some settings, a drug may be restricted to prescription

by selected physicians. However, until followup data are available, the effectiveness of this approach remains unclear. It should be noted that considerations in these doctor-patient discussions might differ according to each woman's risk of becoming pregnant while on the drug.

The third approach involves a formal program of physician and patient education combined, in some cases, with restricted access to the drug. The educational component is intended to assure that physicians and their patients are informed about the drug's teratogenicity and the importance of avoiding pregnancy. The first such effort, in the US, was initiated in late 1988 by the manufacturer of isotretinoin (Accutane[®]), a drug which has a high absolute risk to the fetus and which, at the same time, is uniquely effective in the treatment of severe acne.⁶ Preliminary data from this voluntary "pregnancy prevention program" suggest a high degree of success in achieving both educational objectives, thereby keeping to a minimum the number of malformed infants born as a result of exposure to this agent.⁷ More recently, newly identified uses for thalidomide prompted the US Food and Drug Administration to approve marketing of the drug (as Thalomid[®]) in July 1998, but only with an unprecedented FDA regulated program sponsored by the drug's manufacturer. This program includes an educational component similar to that used for Accutane, but also restricts prescription and distribution of the drug to registered prescribers and pharmacists, respectively. It also mandates that all patients who are prescribed the drug participate in a followup survey designed to monitor and enhance compliance with the manufacturer's "System for Thalidomide Education and Prescribing Safety (STEPS)." The results of these efforts will help determine how best to balance the therapeutic benefits of known human teratogens against the risks of fetal exposure.

DRUGS FOR WHICH TERATOGENIC RISK IS UNKNOWN

For reasons described above, the vast majority of prescription drugs and virtually all nonprescription drugs fall into this category. Regulatory

agencies or manufacturers in various countries may offer a general warning against unnecessary use in pregnancy, but such cautions hardly contribute to rational drug therapy. In settings where the true teratogenic risk is nil, these warnings serve to deny potentially useful drug therapy. Where the true risk is elevated for a particular drug, the nonspecific and "standard" warnings offer little practical discouragement to its use in pregnancy.

DRUGS FOR WHICH TERATOGENESIS IS ALLEGED... AND CLINICAL CONSEQUENCES

At one time or another, a large number of drugs are alleged to be teratogenic, and the clinical consequences can be profound. In one notorious situation, allegations of teratogenicity resulted in a widely used drug being withdrawn from the market. In the late 1970s and early 1980s, the anti-nausea drug Bendectin® (Debendox®, Lenontan®), used widely to treat nausea and vomiting of pregnancy, was alleged to cause a variety of birth defects; the history of this experience has been reviewed elsewhere.⁸ Ironically, the aggregate data on the teratogenic hazards of Bendectin® have ultimately provided the strongest evidence of safety for any medication used in pregnancy. Despite that evidence, however, the manufacturer withdrew the drug from the market as a result of active and potential litigation. At least one study suggests that hospital admissions for hyperemesis gravidarum increased significantly following the drug's withdrawal,⁹ and there is concern that, in the absence of Bendectin, women may be treated with other anti-nausea drugs, although the risk of those drugs to the fetus is unknown.

There are other clinical aspects of unproven allegations, and these are too often ignored. Upon learning that a drug they took in pregnancy might be teratogenic, women who have given birth to malformed infants may become overwhelmed with feelings of guilt. Further, women who are currently pregnant may develop considerable anxiety. That anxiety can lead to a number of clinical consequences, ranging from consultations with physicians to diagnostic procedures (e.g., amniocentesis) to elective termination of the pregnancy. An

experience in the mid-1970s involving a nondrug exposure is instructive. Following widely publicized allegations that spray glue adhesives (typically used to make Christmas decorations) were teratogenic, a US regulatory agency withdrew the product from the market. Although the allegations were subsequently found to be without basis, a survey conducted among genetic counseling centers identified 1100 inquiries from pregnant women prompted by concern about exposure to these agents. Moreover, 11 underwent amniocenteses, and because of this exposure nine women had therapeutic abortions (one had vague evidence of chromosomal damage and eight had no evidence of malformation).¹⁰ These unique and potentially serious clinical consequences argue for heightened attention to scientific rigor and caution in the teratologic assessment of drugs.

THE FALLACY OF "CLASS ACTION" TERATOGENESIS

Another clinically important concern specific to teratogenesis is the issue of "class action". It is widely recognized that an understanding of structure/activity relationships shared by members of a given drug class can be helpful in predicting a given class member's efficacy and adversity (indeed, this view is incorporated into regulatory action in the form of class labeling). However, these predictors cannot be assumed to hold when the adversity at issue is teratogenesis. Given our ignorance about the causes of most birth defects, we cannot know whether it is the chemical structure common to the class that is responsible for teratogenesis or whether the responsible component is that part of the structure that differentiates one class member from another. For example, thalidomide is derived from glutethimide (Doriden® and other brands). Both are glutarimides, and both are sedative/hypnotics. Despite their structural and clinical similarities, thalidomide is clearly a major teratogen and glutethimide is not.¹¹ Thus, we cannot assume that if one drug is a major teratogen, all other members of its class will share that effect. Conversely, we cannot assume that reassurance

about the safety of one drug can be shared by other members of that drug's class.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY RESEARCH

In many ways, the epidemiologic issues involved in the study of birth defects are similar to those of other adverse outcomes; these are considered in detail elsewhere in this text. However, there are a number of considerations that are unique to birth defects or are sufficiently important to warrant particular attention. These have to do with sample size considerations, definitions of exposure and outcome, confounding, and biologic plausibility.

Although serious birth defects occur in approximately three to 4% of liveborn infants, we cannot consider "birth defects" as a single, homogeneous outcome. In fact, physical birth defects include a wide range of malformations that vary in many ways, including their gestational timing, embryologic tissue of origin, and mechanism of development. As examples of variations in timing of occurrence, chromosomal abnormalities generally predate conception; neural tube defects occur in the earliest weeks of gestation; cleft palate occurs toward the end of the first trimester; and microcephaly can occur relatively late in pregnancy. As an illustration of variations in embryologic tissue of origin, some cardiovascular malformations, but not others, are derived from the neural crest cells that migrate from the area surrounding the primitive neural tube. Variations in mechanisms of development include inhibition, disruption, or alteration of the embryologic tissue that is responsible for normal structural development. From a theoretical perspective, then, one would predict that the malformations produced by a drug usually would vary according to the timing of exposure, the sensitivity of the end organ (i.e., embryologic tissue), and the mechanism of its teratogenesis.

Experience supports what would be predicted from biology. Even a brief review of known teratogens reveals a fact that is highly relevant in pharmacoepidemiology studies: most teratogens

do not *uniformly* increase the rates of *all* birth defects, but rather increase rates of *selected* defects. Thus, the "classic" teratogen thalidomide produces defects in about 20–50% of exposed infants, but that increased risk is largely due to increases in limb, spine, and central nervous system malformations.¹² Isotretinoin affects a similar proportion of liveborn infants, but again, that risk is the result of increases in rates of selected specific defects (ear, central nervous system, and cardiac).¹³ Less powerful teratogens behave similarly; valproic acid increases rates of neural tube defects,¹⁴ warfarin increases the rate of cartilage defects,¹⁵ and acetylcholinesterase inhibitors increase rates of renal defects.¹⁶

SAMPLE SIZE CONSIDERATIONS

The fact that pharmacoepidemiology studies must consider rates of *specific* birth defects has a dramatic effect on sample size requirements, both for estimating risk and for providing assurances of safety. With respect to risk in a population of women taking a given drug, a cohort study with a sample size of a few hundred exposed pregnancies might be sufficient to identify a doubling of the three to four percent overall rate of birth defects. Ruling out a doubling of the overall rate would require larger numbers, but these would still be within the same order of magnitude. However, each of the specific defects which together form the overall category of "birth defects" occurs with far less frequency, ranging from about 1 per 1000 livebirths for oral clefts to 1 or fewer per 10 000 for hemimelia/phocomelia. To *detect* a doubling of risk for a relatively common specific birth defect (e.g., 1/1000) would require a sample size of over 20 000 exposed pregnancies (see Appendix Table A.5). To *rule out* a doubling of risk for the same defect, one would need a far larger sample size of exposed pregnancies.

EXPOSURE

There are two concerns regarding drug exposure that require special consideration in birth defects studies. One is the importance of nonprescribed drugs and the other is the issue of recall of drug exposure.

Nonprescribed Drugs

Most epidemiologic research focuses attention on prescribed drugs, in part because many studies utilize data sources with inadequate information on nonprescribed drugs and in part because of the view that prescribed drugs pose the greater potential for risk. As noted earlier, there is no basis for believing that this is the case. Although the effects of nonprescription drugs on the fetus are largely unstudied, it is noteworthy that these agents comprise a substantial proportion of drug exposures in pregnancy.^{17,18} Of potential importance in this category is the increasing use of herbal and other complementary medications whose teratologic effects are completely unknown.

It is therefore important that research include consideration of nonprescribed drugs. In this context, nonprescribed can include not only OTC drugs (including complementary agents), but also prescription drug products that were not intended for the study subject but rather were obtained from friends, neighbors, and relatives: depending on the specific drug, we have found that more than 20% of exposure to prescription drugs can come from such “nonprescribed” sources.¹⁹ Although one might argue about the validity of prescription drug exposure information derived from records (medical, billing, insurance, etc.) versus that derived from patient interviews (see Chapters 13 and 39), there is little question that information on the use of OTC or nonprescribed prescription drugs must be obtained directly from the patient.

Illicit drugs represent a distinct but important subset of nonprescribed drugs. Use of these drugs is seriously under-reported, whether information is drawn from records or interviews. Except in rare settings where exposure may be identified through systematic screening of biologic samples (e.g., urine, blood), epidemiologic studies have major limitations in identifying teratogenic (and confounding) effects of illicit drug use in pregnancy.

Recall Bias

Because of the sample size demands described above, many researchers have turned to the case-control approach for the study of birth

defects. Such studies generally (though not always) rely on maternal interviews for exposure information, and this approach raises concern both about the overall accuracy of recall (see also Chapters 13 and 39) and its susceptibility to bias. More than other drug-induced adverse outcomes, the birth of a malformed child carries an emotional burden and guilt that may affect recall of exposures in pregnancy. When compared to a mother of a normal child, the mother of a malformed infant may be more likely to recall carefully every possible act, event, and drug exposure in pregnancy.²⁰ This tendency is reinforced by repeated inquiries from physicians, nurses, genetic counselors, and relatives, as well as by media and legal attention on the subject of drug-induced birth defects. Thus, in a setting where drug exposure is in fact similar among mothers of normal and malformed infants, one might predict that recall of exposure will be more complete among the latter than among the former. Concern about recall bias is more than theoretical—such bias may well explain a number of drug-defect associations^{21,22} that have subsequently been refuted.^{23,24} On the other hand, evidence supporting the role of recall bias is inconsistent, and the issue of when and to what extent recall bias is present remains an unresolved controversy.

Despite this concern, the simple possibility of recall bias does not invalidate interview-based studies, and there are a number of approaches to reducing and dealing with this problem. These include the choice of controls, the design of the questions, and direct attempts to identify potentially biased recall.

There are differing schools of thought regarding what constitutes an appropriate control for a malformed infant. Some argue that normal infants should be used because of the possibility that a drug might increase the risk of all malformations.²⁵ Since no known teratogen uniformly increases the risk of all malformations, normal controls may be unnecessary. At the same time, use of normal controls might increase the opportunity for differential recall of exposure between mothers of normal and malformed infants. The alternative approach is to use a control group

comprised of infants with a wide range of malformations other than the ones in the case series.²³ By assuring that the controls include a wide range of malformations, one reduces the likelihood that the control series will be biased by inclusion of a large proportion of defects that might be associated with the exposure under study. By restricting comparisons of exposures to those reported by mothers of malformed infants (whether cases or controls), one limits the likelihood of recall bias.²⁶

Both approaches are imperfect. Although no teratogen has yet been identified that uniformly increases the risk of all malformations, the history of teratology is replete with examples of assumptions that proved to be false. About 60 years ago it was thought that the fetus was protected by the placenta from noxious agents, and only 30 years ago it was inconceivable to some that a drug (diethylstilbestrol) could produce cancers in the adult offspring of exposed mothers, or birth defects in the children of women exposed to the drug *in utero*. In an effort to avoid such hubris, some researchers,²⁷ including ourselves, have elected to use two control series, one of malformed and one of normal infants. Since we believe that concern about recall bias exceeds concern about failing to identify an "across-the-board" teratogen, we give primary consideration to findings derived from comparisons with malformed controls.

By definition, recall bias cannot exist if reporting of drug exposure is complete among cases and controls. The closer one comes to that ideal, the less the likelihood of recall bias. It thus becomes critical how one elicits exposure information. Studies that use open-ended questions about drug exposure invite differential recall between mothers of malformed and normal infants.²¹ As might be predicted, the more specifically one asks questions about drug use, the more likely is one to obtain complete information (see also Chapter 39). Recall is also substantially increased when women are asked about use according to various indications, and it is further increased when drugs are asked by specific names.¹⁹ This approach is not likely to result in exaggerated recall (i.e., false positives), as demonstrated by the fact that use of a specific drug

ascertained by such a questionnaire was the same as that estimated from the manufacturer's marketing data; in addition, women did not report exposure when asked about a fictitious drug.²³ In short, by improving ascertainment of drug exposure among both cases and controls, a carefully designed questionnaire can substantially reduce the opportunity for recall bias.

Unfortunately, the possibility of recall bias cannot be completely eliminated either by the use of a malformed control group or by asking specific questions about drug use. In an effort to identify women who might be most at risk for biased recall, we began in 1976 to routinely ask whether a woman had heard that any drug affects the risk of any defect.²³ (This question is asked at the end of the interview, so as not to itself affect reporting of exposures and events.) Our *a priori* assumption is that a woman who acknowledges that a particular drug causes (or prevents) a particular defect is more at risk for differential recall than one who does not. This approach has enabled us to identify indirect evidence of biased recall: in our study of the possible protective effects of folic acid on the development of neural tube defects, we observed different risk estimates when we stratified subjects according to their knowledge of the hypothesis.²⁸ By simply asking women about their perceptions of the teratogenic effects of drugs, one might obtain insight into the nature of biased recall in the study population.

OUTCOME

Given the etiologic heterogeneity of malformations, some have attempted to classify birth defects according to specific categories. We were among those who, more than two decades ago, classified defects by organ system, such as "musculoskeletal" or "cardiovascular."¹ However, classification in this way has little embryologic or teratologic basis, and a more appropriate approach is to create categories that reflect the embryologic tissue of origin. For example, neural crest cells in the earliest stages of embryogenesis migrate to form a variety of structures, including those of the face/ears, parts of the heart, and the neural tube.²⁹ Interference with the normal development of the

neural crest would therefore lead to malformations of tissues derived from neural crest, and that is indeed what has been observed in a number of animal experiments. In fact, these patterns have been observed for certain human teratogens, the most striking example of which is the retinoid isotretinoin, which interferes with neural crest cell migration/development and leads to specific malformations of the ear, heart, and neural tube.¹³ Similarly, certain defects are believed to result from disruption of the embryonic vasculature.³⁰ Although our ignorance about the origins of most birth defects may limit our ability to create categories which share a common etiology, it is preferable, whenever possible, to classify birth defects according to an understanding of their embryologic origins.³¹

CONFOUNDING

As with any other aspect of pharmacoepidemiology research (see Chapter 2), confounding must be taken into account in studies focused on birth defects. Among those variables that require routine consideration are maternal age, race, geography, and socioeconomic status. An understanding of the epidemiology of a given defect or exposure often identifies other variables that may act as confounders in a specific analysis. For example, ethnic background is strongly related to the risk of neural tube defects, maternal age is a strong risk factor for gastroschisis, and alcohol consumption has been associated with defects derived from neural crest. Since medication use may be associated with various other health behaviors (e.g., vitamin use is more common among nonsmokers than smokers), one may need to consider health behaviors, and even nutrition, in studies of certain exposures and outcomes. Further, it may be critically important to separate the teratogenic risk of a drug from the underlying risk associated with the condition for which the drug is taken, something called “confounding by indication”³² (see also Chapters 34 and 43).

Finally, an issue unique to the epidemiologic study of birth defects is the possibility of pregnancy termination. As more malformations be-

come detectable at earlier stages of pregnancy (and as more such pregnancies are terminated), studies of liveborn and stillborn infants will increasingly underestimate the prevalence of such defects. In addition, there are a number of instances where this factor must be considered as a potential confounder (e.g., periconceptional vitamin exposure and neural tube defects).

BIOLOGIC PLAUSIBILITY

Our ignorance about the biologic mechanisms by which most human birth defects occur complicates our ability to determine when a finding may be biologically plausible. There are a few associations for which *in vitro* and animal experiments support their biologic plausibility: these include the increased risk of defects derived from neural crest cells among infants exposed to retinoids, the decreased risk of neural tube defects among infants exposed to folic acid, and the increased risk of defects resulting from vascular disruption among infants exposed to cocaine. However, biologic mechanisms remain unknown for most well accepted drug–defect associations.

In light of this inconsistency, how does one evaluate the importance of biologic plausibility in relation to newly observed associations? On the one hand, a requirement that every association have an identifiable biologic mechanism would have led to dismissal of virtually every accepted human teratogen. On the other hand, some aspects of biologic plausibility must be met. For example, it is implausible that a defect could be caused by an exposure if that exposure first occurs after the gestational development of the defect has been completed. While less absolute, it is unlikely to expect that an exposure would produce a range of defects which span gestational timing from pre-conception to late pregnancy and which do not share embryologic tissue of origin. Thus, we cannot dismiss hypotheses simply because they lack a biologically plausible explanation. However, until they are supported by subsequent studies, such hypotheses must be considered more speculative than hypotheses for which there is a strong biologic basis.

CURRENTLY AVAILABLE SOLUTIONS

There are a variety of approaches used to generate and test hypotheses regarding drugs and birth defects. The purpose of this section is not to list every available data set, but rather to describe the types of resources and their respective strengths and weaknesses. For convenience, these may be divided into cohort and case-control designs. Approaches that involve the monitoring of birth defects without the systematic collection of exposure information are not directly applicable to pharmacoepidemiologic study, and are not considered in this chapter; interested readers are referred to an excellent review.³³

COHORTS

Broadly speaking, there are three types of cohorts relevant to the pharmacoepidemiologic study of birth defects. These are: (i) studies designed to follow large populations exposed to various agents; (ii) the use of data sets created for other purposes; and (iii) followup studies of selected exposures.

Studies Designed to Follow Large Populations Exposed to Various Agents

This approach involves the identification of a population of pregnant women to be followed, with periodic collection of information on demographic characteristics, exposures, and potential confounders, as well as formal evaluation of the offspring at birth and perhaps at some years later as well. A number of studies of this kind have been conducted in various countries.³⁴⁻³⁷ An example is the US Collaborative Perinatal Project (CPP), which enrolled over 58 000 women between 1959 and 1965, obtained detailed information on their pregnancies, and followed the children until age 7.¹ The strength of this type of approach lies in the prospective, systematic, and repeated collection of information that includes exposure to a wide variety of medications taken by a diverse population, many potential confounding variables, and good outcome information.

For commonly used drugs (such as aspirin), there were sufficient data to assess risks for malformations overall as well as for certain subgroups of malformations.³⁸ However, despite the overall size of the database, one major weakness of such a design is the small sample sizes of infants with *specific* malformations. For example, there were approximately 2200 malformed infants in over 50 000 pregnancies followed by the CPP. Among these, there were only 31 with cleft palate (CP) and 11 with tracheoesophageal fistula (TEF). This weakness is further compounded by limited numbers of women exposed to most drugs. If a drug were used by as many as 10% of the women, the expected number of exposed infants with CP and TEF would be three and one, respectively; if a drug were used by 3% of pregnant women, the expected intercepts would be 1 and 0.3. Clearly, these intercepts are so small as to make it infeasible for even so large a cohort to identify any but the most dramatic increases in risk associated with commonly used drugs.

Further, the inordinate costs of such an intensive effort limit enrollment and data collection to a study period of no more than a few years. Because of changing patterns of drug use over time, the clinical relevance of the available data diminishes.

Use of Datasets Created for Other Purposes

In recent years, researchers have focused increasing attention on cohorts identified from databases produced for various purposes (other than for epidemiologic research) by organizations or governments involved in medical care. The strengths and weaknesses vary with the nature of the specific dataset. All have the advantage of identifying exposures prospectively, some may have good reporting of malformations, and some may be derived from large populations. Like most other cohorts, studies based on data from health maintenance organizations (HMOs) may be limited by their small samples of specific malformations. For example, among almost 7000 pregnancies in which 33% of women filled a prescription for Bendectin[®], there were a total of

80 malformations identified, and only 24 affected infants were exposed to the drug *in utero*.³⁹ For the more typical situation where a given drug exposure is far less prevalent, sample size constraints are even more striking: researchers reviewed 15 years of data and identified 215 women who delivered liveborn infants after presumed exposure to topical tretinoin, among whom there were four infants with a variety of malformations.⁴⁰ In both examples, the numbers of exposed infants with specific birth defects were too few to identify even substantial increases in the risk of these outcomes. Further, the definition of exposure was limited to receipt of a prescription for the drug during a given interval, birth defect rates among unexposed subjects were lower than expected, and information on confounding variables was largely absent.

In an effort to overcome sample size limitations, researchers have studied larger datasets. Among these are health insurance data available in the US through the government Medicaid program. In a study of benzodiazepine abuse in pregnancy, researchers reviewed over 100 000 pregnancies and identified 80 women who received numerous prescriptions for the drugs of interest.⁴¹ Unfortunately, records for the offspring could not be found for a substantial proportion of the study subjects, a problem not uncommon in such data sets (see Chapter 19).

Other datasets (and particularly the newer record linkage systems from Scandinavia) offer promise of better information on exposure and outcome variables, but without direct interviews of study subjects they are unlikely to provide reliable information on potential confounding variables (e.g., smoking, alcohol consumption, diet). They are also unlikely to identify exposures to nonprescription drugs, which may be of importance as potential confounders or as primary exposures.

Followup of Selected Exposures

The various mechanisms used to identify cohorts of women exposed to specific drugs include spontaneous reports to manufacturers or government agencies (see Chapters 10 and 11) and registries established by manufacturers and teratogen information services. The strength of such

approaches lies in the ability to identify women exposed to a drug of interest early in pregnancy and, most importantly, to identify and enroll the woman before the outcomes are known. There is the additional advantage of having an opportunity to collect other information prospectively, such as data relating to potential confounding variables. Although the findings from many of these cohorts have been reported in recent years, perhaps the most dramatic—and informative—result was the experience observed among only 36 women who were followed after first-trimester exposure to isotretinoin:¹³ among the 28 liveborn infants, five (18%) were malformed. More striking than the overall rate of malformation was the distribution of defects: each of the five infants had at least one of the specific malformations hypothesized (from premarketing animal studies) to result from isotretinoin exposure (ear, palate, chin, certain heart, and certain brain defects). These findings have been supported by a subsequent study of 94 prospectively identified exposed pregnancies that resulted in livebirths.⁴²

While the above example provides evidence of the value of even small cohorts in detecting dramatic increases in the risk of specific defects, such cohorts are quite limited in their ability to provide assurance of safety. Two examples illustrate this point. In one, the manufacturer followed 276 prospectively identified pregnancies that had been exposed to alprazolam in the first trimester and resulted in a livebirth; 13 malformations were found (4.7%), a rate comparable to that in the general population. The authors pointed out that their sample lacked power to detect even 10–15-fold increases in specific common defects.⁴³ Similar issues surround data drawn from teratogen information services. For example, researchers from four such services identified 128 women who sought counseling because of first trimester exposure to fluoxetine. Among the 98 liveborn infants, there were two with major malformations.⁴⁴

Registries are often limited by problems of self-referral bias and losses to followup (with possible bias based on whether response to followup is related to whether the infant is malformed or normal). In addition, there are limitations inherent

in comparisons between the drug-exposed cohort and the “unexposed.” Some compare observed rates of defects to rates reported in various other populations, and others compare the observed rates to those among pregnancies with exposures to no drugs or to other drugs. Both are imperfect, and comparisons rarely if ever adequately take into account the risk of birth defects due to the condition for which the drug in question was taken.³²

These methodologic concerns aside, all cohort studies have a common limitation. As is clear from the earlier discussion, it is extremely difficult for such a study to achieve sample sizes sufficiently large to provide meaningful reassurance about the safety of a drug. Such reassurance requires information not simply on the overall rate of all birth defects, but also on specific defects, most of which have background rates ranging from 1 per 1000 to 1 per 10 000. Thus, while a cohort of 100 or even 1000 exposed pregnancies might provide reassurance that the drug is not another thalidomide or isotretinoin, such a cohort cannot assure us that a drug is safe with respect to oral clefts, gastroschisis, or other specific birth defects.

CASE–CONTROL STUDIES

The rarity of birth defects in general, and of specific defects in particular, argues for the use of the case–control design in pharmacoepidemiology studies of birth defects. Of course, the strengths and limitations of these studies are similar to those for case–control studies of other outcomes (see Chapter 2), and will not be reviewed here. Such studies may be conducted on an *ad hoc* basis or within the context of case–control surveillance (see Chapter 13). Examples of the latter are few; in North America, two current examples include the longstanding “Birth Defects Study” conducted by our own group,²³ and the more recently established National Birth Defects Prevention Study, involving eight statewide birth defects surveillance programs and coordinated by the US Centers for Disease Control and Prevention.

From the unique perspective of birth defects, case–control studies can have the statistical power required for the assessment of both risk and safety.

At the same time, however, they have the potential limitation of biased recall. There are numerous examples that illustrate both issues, but the following are among the most instructive.

In a cohort study of spermicidal contraceptives, researchers using data from an HMO found the prevalence of birth defects among 763 infants born to exposed mothers to be 2.2% ($n = 17$), whereas the rate among infants born to non-exposed mothers was 1.0%. The excess was attributed to four different defects (including chromosomal defects, limb reduction defects, hypospadias, and neoplasms).⁴⁵ Other investigators used different cohorts to test the hypothesis. Although two analyses involving populations of about 35 000 and 50 000 pregnant women failed to confirm an overall increase in malformation risk,^{46,47} neither study had sufficient power to rule out, with reasonable confidence, an increased risk for each of the specific defects identified in the first study. Therefore, two case–control studies were mounted specifically to test the hypothesis. One identified about 100–400 cases of each of the outcomes of interest, and for a variety of exposure intervals found odds ratios close to unity; more importantly, the upper 95% confidence intervals were 2.2 or lower.⁴⁸ The other identified 151 fetuses with trisomy, including 92 with trisomy 21. Point estimates for various exposure intervals approximated unity, and the study had sufficient power to rule out more than a twofold increase in the risk of trisomy in relation to spermicide use.⁴⁹

While statistical power is a major strength of the case–control approach, power does not assure validity. There are numerous issues that relate to validity (see Chapters 2, 13, and 39). In studies of birth defects, the one that requires particular consideration is recall bias. We previously cited a study of the protective effects of folic acid supplements in relation to neural tube defects, in which we found different risk estimates among women who reported, at the end of the interview, that they were aware of the hypothesis under study.²⁸ We believe that this study supports concern about recall bias. As stated above, however, the simple *possibility* of such bias does not necessarily invalidate a case–control study.

Rather, it requires that investigators consider its existence and make reasonable attempts to identify it in their study population.

We cannot review issues in the epidemiologic study of birth defects without alluding to a concern that cuts across all study designs. Birth defects are complex outcomes, and the study of medications in relation to birth defects only adds to this complexity. For all the reasons described in the introduction to this chapter, the rigorous pharmacoepidemiologic evaluation of birth defects requires considerable understanding and experience not only of epidemiology, but also of related disciplines (e.g., pharmacology, embryology, teratology).

AN INTEGRATED APPROACH

From a pharmacoepidemiologic perspective, there are two broad concerns regarding teratogenesis. The first, both in theory and in practice, is to identify major teratogens (exemplified by thalidomide and isotretinoin). The second is to identify teratogens with more modest risks (exemplified by phenytoin and valproic acid). Although not widely recognized, a combination of the approaches described above can provide much of the information needed to respond to both concerns. Cohorts of exposed subjects, whether in the form of registries or record linkage studies, can identify major teratogens in a relatively timely way. In most instances, the extremely large risks associated with drugs such as thalidomide or isotretinoin tend to overwhelm the distinct methodologic limitations inherent in these approaches. If a drug makes it past that first line of defense, case-control surveillance (or focused case-control studies) can provide the power and rigor necessary to identify more modest teratogens. Simple sample size considerations will continue to limit our ability to demonstrate safety with respect to relatively rare exposures or outcomes, but an integrated approach that combines cohort and case-control studies offers an effective step towards resolving these two teratogenic concerns.

THE FUTURE

Studies of birth defects in the future will undoubtedly focus increased attention on issues of statistical power, validity, and secular changes in exposures. The recent approval of thalidomide in the US will, one hopes, focus renewed attention on the teratogenic potential of drugs and how best to identify it. While the effectiveness of newer regulatory approaches (e.g., isotretinoin and thalidomide) remains to be determined, there appears to be growing acceptance of the notion that proven human teratogens may require a unique form of regulatory attention and control if they are to be used with relative safety.

There are two issues that warrant particular attention. These are the increasing integration of epidemiology and biology and the legal climate in which epidemiologists will operate.

INTEGRATION OF EPIDEMIOLOGY AND BIOLOGY

A major frustration among those who conduct epidemiologic studies of congenital malformations is the dearth of understanding of the mechanisms, both structural and molecular, by which defects occur. Rapid advances (such as those related to retinoic acid) will markedly enhance our ability to classify defects in biologically meaningful categories. Advances can also be expected in the identification of drug exposures in human tissue. Blood and urine have long been available for this purpose, but detection is largely limited to the interval shortly following exposure. For case-control studies in particular, where the mother is typically identified some time after delivery, such sampling is of no use for detecting exposures in early pregnancy. In recent years, however, researchers have explored the usefulness of other tissues in which drugs or their metabolites may persist and accumulate, such as meconium⁵⁰ and hair.⁵¹ Although these techniques are still under evaluation and currently lack the ability to estimate timing precisely, they may enable researchers at least to confirm the presence or absence of certain exposures during pregnancy.

With respect to the role of drugs in the etiology of birth defects, there is no question that the most exciting biologic development is the rapidly expanding field focused on genetic markers of susceptibility to teratogenesis. It has puzzled many that known human teratogens do not produce malformations in all (or even most) exposed fetuses. It was widely believed that this phenomenon was due to differences in host susceptibilities, and in 1985 researchers demonstrated that such a phenomenon was likely to account for the inconsistent effect of at least one such teratogen—phenytoin.⁵² These workers found that a genetic defect in the detoxification of arene oxide (a radical metabolite of phenytoin) was strongly related to the risk of the major defects associated with phenytoin in numerous studies. Such genetically determined variations in drug metabolism (“genetic polymorphisms”) have been described for other drugs (such as sulfonamides and alcohol), and no doubt more will be identified in the future. In anticipation, we and others are adding databanks of buccal cells or blood samples to ongoing studies of risk factors for birth defects. An understanding of genetic polymorphisms will dramatically enhance the identification of subsets of the population who are at increased risk for certain birth defects and the identification of drugs that might warrant particular study.

By analogy with the process of screening for rubella susceptibility or for genetic diseases, it is not unreasonable to look forward to a time when women of childbearing age can be screened for drug-specific genetic polymorphisms that place them at particular risk for having a malformed infant. Information of this kind has obvious usefulness in selecting (and avoiding) specific drugs for the treatment of women who are pregnant or at risk for becoming pregnant.

THE LEGAL CLIMATE

Whatever their design, studies of exposures in pregnancy in relation to birth defects ultimately depend on the ability to link exposure and outcome information. To accomplish such linkage, researchers require access to information which identifies women who have become pregnant, the

outcomes of those pregnancies (including spontaneous and therapeutic abortions), and details as to the presence or absence of malformations. The issue of whether and how such information might be disclosed to researchers has become highly contentious in many countries (see also Chapter 26). For case-control studies in particular, the enrollment of malformed and/or normal subjects requires that hospitals, other health providers, or government agencies make identifying information available to researchers, who then contact eligible subjects in order to invite them to participate in an interview. At present, there is considerable public anxiety—and even anger—regarding the erosion of confidentiality, especially with respect to financial and related data. While adults with cancer or heart disease are unlikely to view being approached to participate in research as an invasion of privacy, parents of children with birth defects, and particularly those who have undergone therapeutic abortions, are exquisitely sensitive to the disclosure of information on their pregnancy and its outcome.

We are unaware of any evidence to suggest that medical researchers have compromised confidentiality. Nonetheless, there is a real possibility that epidemiologic research into drug-induced birth defects may be constrained or even eliminated by actions and laws intended to protect confidentiality, without consideration to the substantially different needs filled by medical research and commercial interests.⁵³ It is therefore critical that the public be educated about the extent to which epidemiologic research serves the public health, and that they recognize that this benefit can only be accomplished by the provision of confidential medical information to legitimate researchers. At the same time, the public must be reassured—and researchers must accept—that violations of this shared trust will be accompanied by serious penalties.

ACKNOWLEDGEMENTS

I wish to express my appreciation and thanks to Samuel Shapiro, MB, MRCP(E), Carol Louik, Sc.D., and Martha Werler, Sc.D., whose advice and

close working relationships have contributed to the issues covered and views expressed in this chapter.

This work was supported by NIH Grants HD27697 and HL58763; unrestricted additional support to the Birth Defects Study of the SEU has been provided by Rhone Poulenc Rorer Pharmaceuticals, Inc., Hoechst Marion Roussel, Inc., Glaxo-Wellcome Inc., and Pfizer Inc.

REFERENCES

- 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, 1977; 31.
- CDC. *Congenital Malformations Surveillance Report: January 1982–December 1985*. Atlanta: US Department of Health and Human Services, Public Health Service, 1988.
- Gregg NM. Congenital cataract following german measles in the mother. *Trans Ophthalmol Soc Aust* 1941; **3**: 35–46.
- Lenz W. Thalidomide and congenital abnormalities. *Lancet* 1962; **1**: 45.
- Warkany J. Problems in applying teratologic observations in animals to man. *Pediatrics* 1974; **53**: 820.
- Stern RS. When a uniquely effective drug is teratogenic: the case of isotretinoin. *New Engl J Med* 1989; **320**: 1007–9.
- Mitchell AA, Louik C, Van Bennekom C. Can human teratogens be used with relative safety? A novel survey method. Presented at American Epidemiologic Society, Pittsburgh, PA, 1993.
- Holmes LB. Teratogen update: Bendectin. *Teratology* 1983; **27**: 277–81.
- Neutel CI, Johansen HL. Measuring drug effectiveness by default: the case of Bendectin. *Can J Public Health* 1995; **86**: 66–70.
- Hook EB, Healy KB. Consequences of a nationwide ban on spray adhesives alleged to be human teratogens and mutagens. *Science* 1976; **191**: 566–7.
- Heinonen OP, Slone D, Shapiro S. Sedatives, tranquilizers, and antidepressant drugs. In: Kaufman DW, ed., *Birth Defects and Drugs in Pregnancy*. Littleton, MA: Publishing Sciences, 1977; 343–4.
- Newman CGH. Teratogen update: Clinical aspects of thalidomide embryopathy: a continuing preoccupation. *Teratology* 1985; **32**: 133–44.
- Lammer EJ, Chen DT, Hoar RM, et al. Retinoic acid embryopathy. *New Engl J Med* 1985; **313**: 837–41.
- Robert E. Valproic acid as a human teratogen. *Cong Anom* 1988; **28** (suppl): S71–80.
- Whitfield MF. Chondrodysplasia punctata after warfarin in early pregnancy. *Arch Dis Child* 1980; **55**: 139–42.
- Martin RA, Jones KL, Mendoza A, Barr M, Benirschke K. Effect of ACE inhibition on the fetal kidney: decreased renal blood flow. *Teratology* 1992; **46**: 317–21.
- Werler MM, Mitchell AA, Shapiro S. First trimester maternal medication use in relation to gastroschisis. *Teratology* 1992; **45**: 361–7.
- Rayburn W, Wible-Kant J, Bledsoe P. Changing trends in drug use during pregnancy. *J Reprod Med* 1982; **27**: 569–75.
- Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.
- Werler MM, Pober BR, Nelson K, Holmes LB. Reporting accuracy among mothers of malformed and nonmalformed infants. *Am J Epidemiol* 1989; **129**: 415–21.
- Rothman KJ, Fyler DC, Goldblatt A, Kreidberg MB. Exogenous hormones and other exposures of children with congenital heart disease. *Am J Epidemiol* 1979; **109**: 433–9.
- Bingol N, Fuchs M, Diaz V, Stone RK, Gromisch DS. Teratogenicity of cocaine in humans. *J Pediatr* 1987; **110**: 93–6.
- Mitchell AA, Rosenberg L, Shapiro S, Slone D. Birth defects related to Bendectin use in pregnancy: 1. Oral clefts and cardiac defects. *J Am Med Assoc* 1981; **245**: 2311–4.
- Slutsker L. Risks associated with cocaine use during pregnancy. *Obstet Gynecol* 1992; **79**: 778–89.
- Bracken MB. Methodologic issues in the epidemiological investigation of drug-induced congenital malformations. In: Bracken MB, ed., *Perinatal Epidemiology*. New York: Oxford University Press, 1984; 435–40.
- Lieff S, Olshan AF, Werler M, Savitz DA, Mitchell AA. Selection bias and the use of controls with malformations in case-control studies of birth defects. *Epidemiology* 1999; **10**: 238–41.
- Mills JL, Rhoads GG, Simpson JL, et al. The absence of a relation between the periconceptional use of vitamins and neural tube defects. *New Engl J Med* 1989; **321**: 430–5.
- Werler MM, Shapiro S, Mitchell AA. Periconceptional folic acid exposure and risk of occurrent neural tube defects. *J Am Med Assoc* 1993; **269**: 1257–61.
- Johnston MC, Sulik KK. Some abnormal patterns of development in the craniofacial region. *Birth Defects Orig Artic Ser* 1979; **15**: 23–42.
- Hoyme HE, Jones MC, Jones KL. Gastroschisis: abdominal wall disruption secondary to early gestational interruption of the omphalomesenteric artery. *Semin Perinatol* 1983; **7**: 294–8.

31. Khoury MJ, Moore CA, James LM, Cordero JF. The interaction between dysmorphology and epidemiology: methodologic issues of grouping and splitting. *Teratology* 1992; **45**: 133–8.
32. Slone D, Shapiro S, Miettinen OS, *et al.* Drug evaluation after marketing. *Ann Intern Med* 1979; **90**: 257–61.
33. Khoury MJ, Holtzman. On the ability of birth defects monitoring to detect new teratogens. *Am J Epidemiol* 1987; **126**: 136–43.
34. Michaelis J, Gluck E, Michaelis H, Koller S, Degenhardt K-H. Teratogene Effekte von Lenotan? *Dtsch Arztebl* 1980; **23**: 1527–9.
35. Gibson GT, Colley DP, McMichael AJ, Hartshorne JM. Congenital anomalies in relation to the use of doxylamine/dicyclomine and other antenatal factors. *Med J Aust* 1981; **1**: 410–4.
36. Harlap S, Shiono PH, Ramcharan S, *et al.* Chromosomal abnormalities in the Kaiser-Permanente Birth Defects Study, with special reference to contraceptive use around the time of conception. *Teratology* 1985; **31**: 381–7.
37. Czeizel A. Lack of evidence of teratogenicity of benzodiazepine drugs in Hungary. *Reprod Toxicol* 1988; **1**: 183–8.
38. Slone D, Siskind V, Heinonen OP, *et al.* Aspirin and congenital malformations. *Lancet* 1976; **1**: 1373–5.
39. Jick H, Holmes LB, Hunter JR, *et al.* First-trimester drug use and congenital disorders. *J Am Med Assoc* 1981; **246**: 343–6.
40. Jick SS, Terris BZ, Jick H. First trimester topical tretinoin and congenital disorders. *Lancet* 1993; **341**: 1181–2.
41. Bergman U, Rosa FW, Baum C, Wiholm B-E, Faich GA. Effects of exposure to benzodiazepine during fetal life. *Lancet* 1992; **340**: 694–6.
42. Dai WS, LaBraico JM, Stern RS. Epidemiology of isotretinoin exposure during pregnancy. *J Am Acad Dermatol* 1992; **26**: 599–606.
43. St. Clair SM, Schirmer RG. First-trimester exposure to alprazolam. *Obstet Gynecol* 1992; **80**: 843–6.
44. Pastuszak A, Schick-Boschetto B, Zuber C, *et al.* Pregnancy outcome following first-trimester exposure to fluoxetine (Prozac). *J Am Med Assoc* 1993; **269**: 2246–8.
45. Jick H, Walker AM, Rothman KJ, *et al.* Vaginal spermicides and congenital disorders. *J Am Med Assoc* 1981; **245**: 1329–32.
46. Mills JL, Harley EE, Reed GF, Berendes HW. Are spermicides teratogenic? *J Am Med Assoc* 1982; **248**: 2148–51.
47. Shapiro S, Slone D, Heinonen OP, *et al.* Birth defects and vaginal spermicides. *J Am Med Assoc* 1982; **247**: 2381–4.
48. Louik C, Mitchell AA, Werler MM, Hanson JW, Shapiro S. Maternal exposure to spermicides in relation to certain birth defects. *New Engl J Med* 1987; **317**: 474–8.
49. Warburton D, Neugut RH, Lustenberger A, Nicholas A, Kline J. Lack of association between spermicide use and trisomy. *New Engl J Med* 1987; **317**: 478–82.
50. Ostrea EM Jr, Romero A, Yee H. Adaptation of the meconium drug test for mass screening. *J Pediatr* 1993; **122**: 152–4.
51. Graham K, Koren G, Klein J, Schneiderman J, Greenwald M. Determination of gestational cocaine exposure by hair analysis. *J Am Med Assoc* 1989; **262**: 3328–30.
52. Strickler SM, Dansky LV, Miller MA, *et al.* Genetic predisposition to phenytoin-induced birth defects. *Lancet* 1985; **2**: 746–9.
53. International Society for Pharmacoepidemiology. *Data Privacy, Medical Record Confidentiality, and Research in the Interest of Public Health*. 1997, amended 1998.

43

Bias and Confounding in Pharmacoepidemiology

JEAN-PAUL COLLET and JEAN-FRANÇOIS BOIVIN

Department of Epidemiology and Biostatistics, McGill University, Montréal, Canada
Centre for Clinical Epidemiology and Community Studies, SMBD Jewish General Hospital, Montréal, 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* itself is the absence of both random and systematic error. The question of measurement has always represented a key point in epidemiology; Rothman wrote that “an

epidemiologic study is properly viewed as an exercise in measurement, with accuracy as the goal.”¹

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLGY RESEARCH

In 1981 Alderslade and Miller presented the results of the National Childhood Encephalopathy Study (NCES), a nationwide case-control study conducted in the UK initiated to answer the question of a possible association between diphtheria-tetanus-pertussis (DTP) vaccine and the subsequent development of neurologic disorders.² This report received tremendous publicity because of the importance of the question, but it also raised a large international controversy because several potential biases affected the credibility of the results.

In the NCES, physicians of England, Scotland, and Wales reported 1182 cases of severe acute neurologic illnesses in infants and children aged 2–35 months; two controls were selected for each case, matched on age, gender, and residential area. The NCES found that the risk of severe acute neurologic event was significantly increased within the seven days following DTP vaccine (relative risk (RR) = 2.3; 95% confidence interval = 1.4–3.2), and that, one year after the vaccine, seven 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 three of 478 controls (0.6%), yielding a relative risk of 4.7 (95% confidence interval = 1.1–28.0). The results have been used in many court trials by parents of disabled children seeking compensation.

It was, however, the beginning of an important controversy. First, the NCES's results were not confirmed in several other, smaller, studies.^{3–6} In addition, several biases were considered possible explanations, partial or total, for what was observed. The following problems, in particular, were discussed.

- The fact that the participating physicians knew the objectives of the study may have increased the reporting of cases that occurred shortly after vaccination. This referral bias would have increased the apparent relative risk.
- Information bias was also considered possible. The exact date of onset of the neurological problem was occasionally difficult to ascertain precisely. Also, the interviewers and data collectors were not blinded to the study's objectives, nor to the participants' clinical status. It is thus possible that the date of onset was sometimes shifted toward a shorter post-vaccine time window. The fact that the immediate high risk post-vaccine period was followed by a subsequent low risk period has been considered as evidence that this bias actually occurred.
- The main concern was selection bias. Children who had previous neurologic disorders (e.g., seizures) were more likely to go unvaccinated and were also more likely to develop permanent brain damage. This selection bias would have favored the inclusion of nonvaccinated individuals in the case group, reducing the apparent relative risk. It is also possible that the previous neurologic disorder was present and occasionally the case "event" was the first expression of the disease. In that case, these children should have been excluded from the study because the disease in fact preceded the exposure (see "protopathic bias" below). Of course, in this situation the subclinical neurologic disease could not have biased the choice of whether or not to vaccinate.
- Several other issues were also of concern in the interpretation of the NCES results. These include, for example, the lack of precision of the disease definitions as well as the inclusion of cases which were not thought to be plausibly related to DTP vaccine (such as Reye's syndrome, hypsarrhythmia, or acute viral encephalopathies), with a possible dilution of any vaccine-specific association. The fact that the odds ratio for the risk of permanent brain damage was based on only seven cases and three controls was of particular concern, because among the seven cases there was one case of Reye's syndrome and two cases of well identified viral encephalitis.

The NCES represents an example of how pharmacoepidemiology studies may contribute to data needed by clinicians, public health administrators, and lawyers. The controversy surrounding it, however, shows that a well thought out and designed pharmacoepidemiology study can raise many questions and lose an important part of its credibility because of the possibility of biases. Once a study is completed, some biases can only be discussed in term of likelihood and importance, and cannot be estimated precisely, nor controlled at the analysis stage. The possibility of biases in a study raises doubts about its results, and provides an opportunity for subsequent debate based upon convictions rather than evidence.

Generally speaking, questions raised about the quality of studies, as exemplified above, have important implications for clinicians, regulators, public health administrators, and health economists. For example, when a case-control study shows that the risk of a severe adverse event is different in patients treated, respectively, with drugs A and B, the following questions should be raised.

- Was the clinical indication that led to the prescription of drug A or B considered in the analysis? Different drugs are likely to be prescribed to different types of patient, with different diseases or different degrees of severity or clinical expression of the same disease; such differences in baseline characteristics may lead to the patients being at different risks of developing an adverse outcome, independently of the prescription of drug A or B.
- Were all eligible cases included? Was the exclusion of some patients related to drug exposure?
- How precisely was exposure measured? Exposure is often reduced to a dichotomous classification, which is not very precise.
- Could previous knowledge of the problem have influenced the reporting of cases or the recall of previous exposure?
- Was the exposure time window defined in relation to the onset of outcome or in relation to the onset of drug exposure?
- Was temporal information about drug exposure included in the analysis? This information is very important because the risk is probably very different in new users, in long term chronic users of the same drug, and in those who have shifted from one drug to another.
- Was the strength of the association constant in all categories of patients (i.e., independent of age, co-morbidity, co-prescription, different forms of the disease, etc.)?

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 (e.g.,^{1,7-11}), and discussed in some important articles.¹²⁻¹⁶ Pharmacoepidemiology studies, however, may be affected more often by some particular biases than other epidemiologic studies. Moreover, the dynamic 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 category from other types of bias (especially selection bias). We will also discuss the dynamics of bias in pharmacoepidemiology and finally, 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 PHARMACOEPIDEMOLOGIC RESEARCH

BIASES IN PHARMACOEPIDEMOLOGY

Several potential biases are likely to affect pharmacoepidemiology studies, such as referral bias, recall bias, nondifferential or differential follow-up bias, protopathic bias, and prevalence study bias. For the purpose of our discussion, however, we will classify these biases into three groups: selection bias, information bias, and confounding. Figure 43.1 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 also covariates such as confounding variables or effect modifiers. Confounding is related to the pathophysiological mechanisms of disease development, which may be affected by several factors acting together; what is observed is due not to the exposure of interest, but to other factors. Confounding variables may also represent risk factors

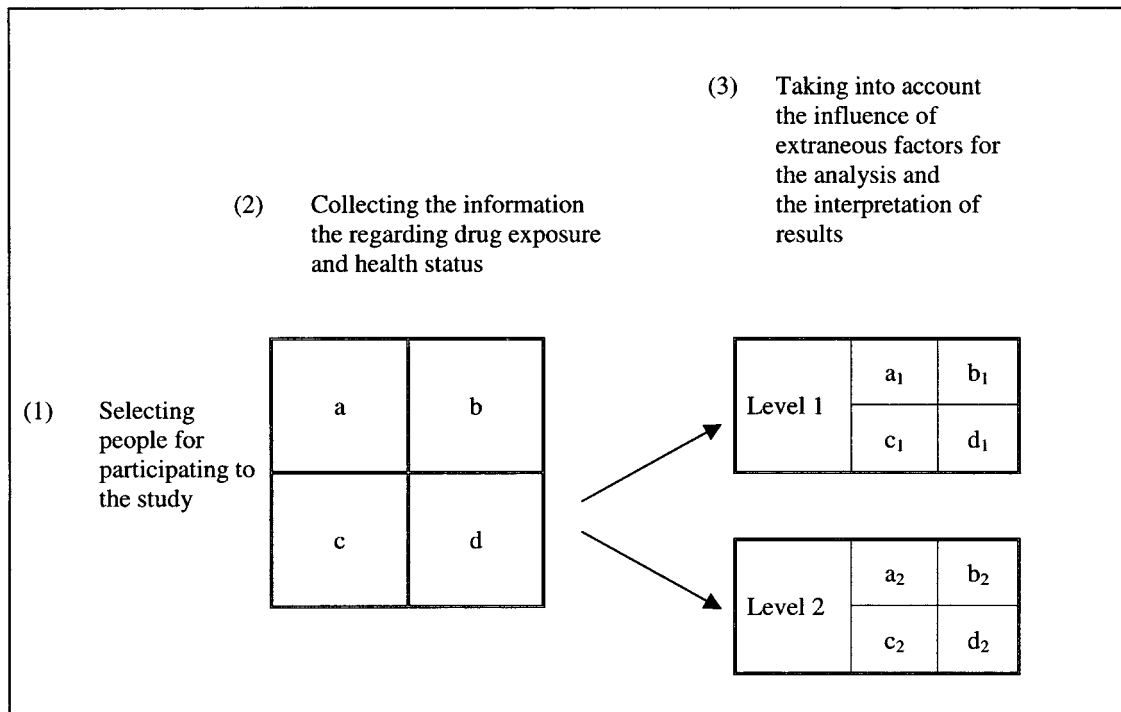


Figure 43.1. Biases in pharmacoepidemiology.

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.

a, disease +, exposure +; b, disease -, exposure +; c, disease +, exposure -; d, disease -, exposure -.

for the outcome (for example, age), without being causal factors.

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 have an unusual and unequal relationship between drug exposure and outcome. Hence, the estimate of the association in the study group differs from the estimate 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, and prevalence study bias, with the special case of "protopathic bias."

Referral Bias

Referral bias can occur if the reasons for referring a patient, to the hospital for instance, may be 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, if one were studying serious gastrointestinal bleeding, this might not be a problem.

The same problem may occur with diagnoses of deep venous thrombosis in women using oral contraceptives who present with leg pain. Knowledge of a well established association between the disease and the drug makes the use of an oral contraceptive a key element in the diagnosis, such that these women may be more likely to be subjected to diagnostic tests for venous thrombosis.

This referral bias problem may even be more generalized. It is possible, for instance, that in older people taking many drugs, the occurrence of "any change in health status" or "any suspicious side effect" may lead to hospitalization just because of exposure to a large number of drugs. In these circumstances, there will be a positive association between drug use in general and any health condition.

Appreciating the potential for referral bias helps in interpreting the results of successive epidemiologic studies conducted at different points in time. We can imagine a situation in which only the first study was unbiased because the association was not yet known. When the first publication shows a positive association between drug and disease, the referral bias phenomenon begins and is likely to increase after each new report of a positive association. In this context, we may observe an increase in the strength of the association over time, even if the true association remains constant and even if it is actually null.

This phenomenon of referring "exposed cases" may occasionally have a very large impact in pharmacoepidemiology. It has been shown, for example, that the publication of a letter in a medical journal may affect a physician's ability to find new cases of the condition and increase the probability of his or her reporting them. This type of alert is also likely to increase the referral of these "interesting" patients to the hospital; hospitalized patients will then represent a biased group with a strong positive association between drug exposure and health event. Reports on side effects of triazolam, for example, are likely to have been

affected by referral bias: after the first publication,¹⁷ several authors reported other single case experiences,^{18,19} and then Sunter reported a series of cases in 1988.²⁰ While these were all spontaneous reports of adverse reactions, not formal epidemiologic studies, one could speculate that subsequent formal studies could have been affected by referral bias.

One mechanism to prevent this type of *publicity bias* would consist of keeping confidential the reports of single cases of possible adverse events until a good epidemiologic study has confirmed them. From this perspective, confidentiality related to the publication of a single case report should not be viewed as a protection of the manufacturer but as a protection of the scientific truth. This is because once the first case report has been "mediatized," the situation is intrinsically biased and, in some situations, the question can no longer be answered, or can only be answered with a great deal of difficulty. Of course, whether such a confidential treatment of case reports would be appropriate from a public health perspective can be debated. 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, all serious cases will eventually be diagnosed correctly.

Self-selection

When patients decide themselves to participate or to leave a study, a type of selection bias may occur, because this decision may be related both to drug exposure and to 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 interview-based historical case-control studies, because both outcome and exposure are already manifest when study subjects are recruited. It could be the case, for instance, when studying birth defects; 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. Such a situation is not unusual in pharmacoepidemiology. The problem must be controlled at the design level

by incorporating systematically all the cases suffering from the condition. Having a registry is an excellent way to cope with this problem of selection. A similar problem may occur in historical cohort studies relying on volunteers. Losses to followup may also bias the results similarly in cohort studies, if those who drop out belong to a special disease–exposure category.

Prevalence Bias

Another type of selection bias may occur in case–control studies when prevalent cases rather than new cases of a condition are selected for a study. An 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. 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 positive association with “good prognosis prevalent cases” might not be confirmed in the whole group of patients defined by incidence. The situation could even be worse if only one group of patients defined by their disease duration receives the medication. It would be the case, for example, if only survivors receive the medication; there will obviously be a positive association between drug use and outcome.

Another similar bias may occur when patients are recruited through a screening procedure. The population selected by screening may be different from the one identified by clinically manifest symptoms. There may be different relationships between drug exposure and adverse effects in these two subpopulations.

Generally speaking, criteria for inclusion, even those not clearly stated, must be considered with the highest attention when interpreting the results from nonexperimental studies; preventing selection bias or dealing with it is an essential stage in the development of a protocol in pharmacoepidemiology.

“Protopathic Bias”

The term “*protopathic bias*” was first used by Feinstein.⁸ It may occur “if a particular maneuver

was started, stopped, or otherwise changed because of the baseline manifestation caused by a disease or other outcome event...” Confusion between cause and effect may then arise. 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 find a negative association between current aspirin use and colon cancer. This type of bias is a consequence of selecting only clinically manifest cases, and of the difficulty of precisely ascertaining exposure that occurred in the past. This situation may occur in pharmacoepidemiology because diseases are often identified late after their first clinical expression and exposure to drugs may change from day to day. Such possibilities demonstrate the paramount importance of as full an understanding as possible of the pathophysiologic mechanism of disease development in designing pharmacoepidemiologic 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 sick people may be considered normal (and the reverse). 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*, which means systematically biased. 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 Misclassification

When the degree of misclassification is similar for all patients and independent of both exposure and health status conditions (because the instrument is not very reliable, for instance), it is called nondifferential or random misclassification. 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. This tendency is very important to consider when the conclusion of the study is that there is “no statistically significant association.” This may merely be due to important random misclassification. In pharmacoepidemiology, the assessment of drug exposure is likely to be affected by an important rate of nondifferential misclassification because it is related to many factors that are

difficult to control. Figure 43.2 shows the different possibilities for misclassifying drug exposure. This is also discussed in more detail in Chapter 39.

Another important misclassification may occur when dichotomizing a patient’s exposure as “exposed/not exposed” without taking into account the timing of the exposure. The International Agranulocytosis and Aplastic Anemia Study (IAAAS)²¹ was criticized because the authors defined their time window only in relation to the occurrence of outcome and not in relation to the onset of drug exposure, ignoring patients’ history of previous exposure.²² The fact that previous, especially long term, exposure to the drug was likely to be associated with a lower risk of severe events could have distorted the results. This lack of accuracy in defining exposure may result in an important information bias, which may lead to a nonsignificant association overall while, in fact, within a specific time window, there is a very strong association between the drug and the outcome.

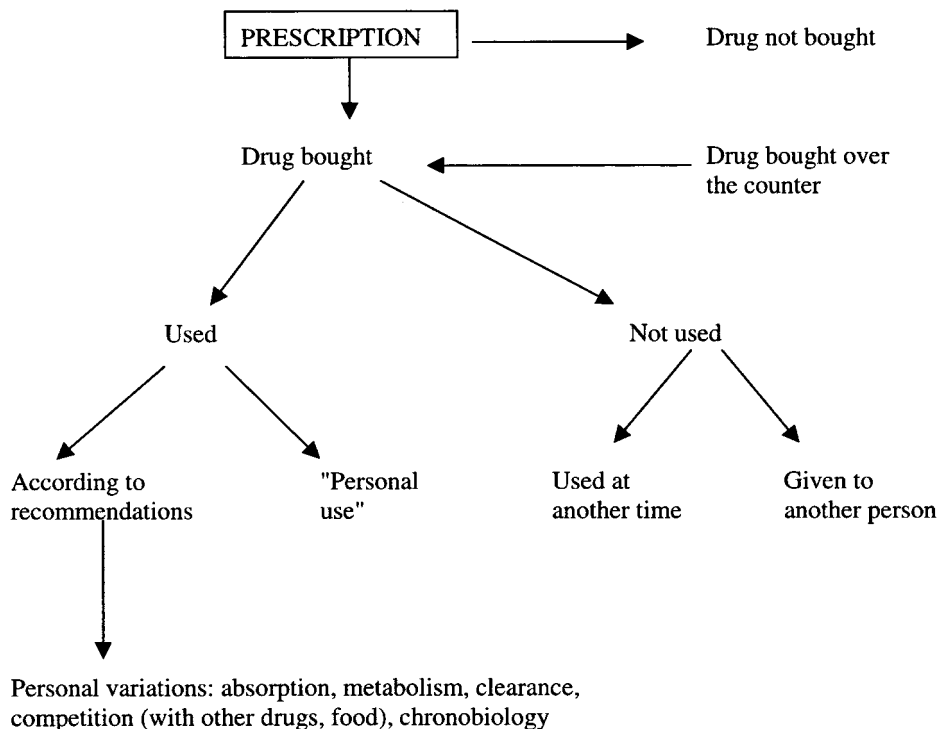


Figure 43.2. Factors influencing drug exposure.

Figure 43.3 shows different hazard functions related to drug exposure. Anaphylactic reactions occur rapidly after drug exposure, the risk being very high during a short period of time and null after this 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. 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). In the example shown in Figure 43.3, the instantaneous risk is completely different in each situation but the average risk for the full period is similar. The

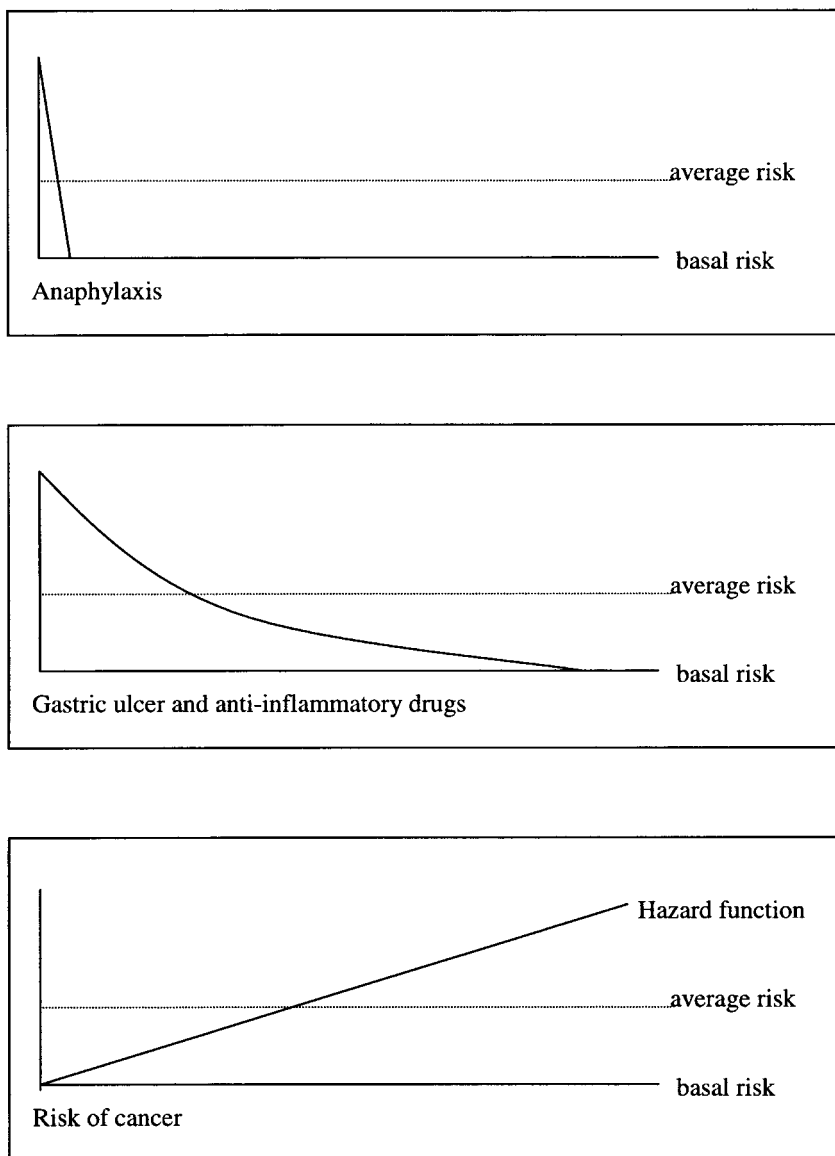


Figure 43.3. Hazard functions related to drug exposure.

average risk for the total period of follow-up does not represent the real risk faced by the patients. This mismeasurement is likely to decrease the strength of the association between drug exposure and adverse event.

Differential Misclassification

When misclassification is related to the exposure–outcome association, we refer to it as differential or systematic misclassification. This possibility occurs each time the knowledge about the outcome status in a case–control study, or the exposure status in a cohort study, influences the validity of the information collected. 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 different 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. This has nothing to do with deliberate lies or a desire to mislead, but to memory being better when there is good reason for it. This problem must be controlled at the design level by choosing, for instance, a control group likely to have the same memory of the past, if possible, e.g., alternative birth defects (see also Chapter 42).
- *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., more attention is given to assessing exposure in the 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. Differential follow-up may sometimes be very subtle. For instance, if a drug is responsible for the development of side-effects that require specific care (e.g., abdominal pain requiring radiologic examination or medical visits), the differential follow-up induced by this condition may be responsible for an excess of diagnoses of several other already prevalent conditions that would otherwise have not been identified (e.g., cholelithiasis).

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 other variables that are also risk factors for the outcome of interest. Confounding occurs when the distribution of these risk factors is unbalanced across the different levels of the drug exposure. In this case, the occurrence of the outcome is changing from one level of the drug exposure to the other in relation (partly or totally) to the cofactors. Without good information on the other risk factors, i.e., valid and precise measurement of their magnitude, it is not possible to separate the respective effect of each component. The estimate measured in this condition is said to be “confounded” by the other cofactors.

In this section of the chapter, we shall present some numerical examples aimed at showing the mechanism of confounding and the statistical principles for correcting its effect. We shall also describe some important confounders in pharmacoepidemiology and we shall show that it is sometimes difficult to distinguish 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 43.4 shows the classical relationship between the three variables. For example, when studying the relationship between the use of nonsteroidal 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.

In such a situation, the measure of association between nonsteroidal anti-inflammatory drugs and gastric ulcers can be affected by the confounder. In the extreme situation, physicians could decide never to use nonsteroidal anti-inflammatory drugs in patients with a history of gastric problems. These drugs could then appear as protective for peptic ulcer. The result in this case would obviously be biased. A confounder may be responsible for a part or all of the effect observed. It may exaggerate, mitigate, or reverse a true effect. The existence of confounding is generally established by comparing the crude estimate of the

association with the estimate obtained after controlling for the potential confounder.

Table 43.1 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 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

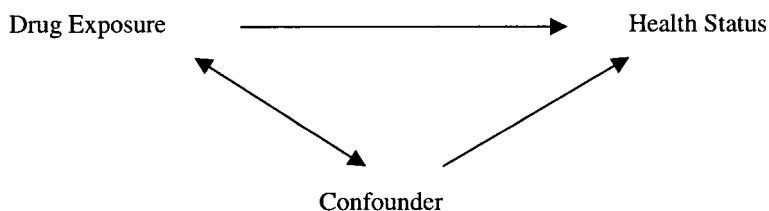


Figure 43.4. Mechanism of confounding.

Table 43.1. 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	

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 43.1 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 43.2 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 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. In the previous example, there was a reversal of effect.

Table 43.2 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. A statistical test is usually used to assess whether such variation between strata can be attributed to random fluctuations, or whether it represents a true effect.

Important Confounders in Pharmacoepidemiology

Confounding by the Reason for Prescription

The indication for a prescription is probably the most important confounding factor in pharmacoepidemiology since, theoretically, there is always a reason for prescription and because the reason is often associated with the outcome of interest. The situation has been the object of important considerations in the discussion of study results^{23–26} (see also Chapter 34). It has also received several different names such as “indication bias,” “channelling,” or “confounding by severity.” All these labels, as well as others such as “contra-indication bias,” just represent the fact that there is a reason for prescribing a drug.

The problem of confounding by indication can be considered in the same perspective as selection bias, because the decision to prescribe can be viewed as one way to select 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 34), there is a bias. This perspective shows that, for a given drug, the

Table 43.2. 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	1110	
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	

possibility of bias is not universal, but is directly related to the outcome studied and may also change over time, or from one country to another. It also shows the extreme difficulty of adjusting for this type of confounding. Miettinen has discussed this problem in the context of a study of intended effects:²⁴ "Thus, perceived high risk or poor prognosis tends to constitute an indication for intervention.... Where such an indication does guide the use of intervention, it constitutes a confounder: it is, by definition, a correlate of the determinant at issue, the intervention, and it is a predictor of the null outcome in so far as it indeed does imply the presumed high risk or poor prognosis."

The term "confounding by indication" implies that we could control for the reason for the prescription at the analysis stage. Although this is theoretically possible, it is in practice 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 very complex and multifactorial phenomenon involving the physician's knowledge and many factors, sometimes not rational, which act in different directions. Miettinen 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.²⁴ 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. Miettinen showed further in that example that controlling for the information available related to the reasons for the prescription, e.g., previous history of blood clotting, reduced the bias but could not change the direction of the association: after adjustment the risk in treated patients was still four times higher than in nontreated patients.

The example illustrates that it can be very difficult to measure accurately the reasons for prescribing. This rather pessimistic conclusion has two consequences. The first one is related to the scope of research in pharmacoepidemiology, in which the

study of the determinants of prescribing is a domain requiring much attention. The second is related to the desirability of randomized clinical trials, rather than nonexperimental studies, whenever the results from nonexperimental studies are likely to be biased and "inconclusive" because of confounding by the indication. A more positive view of this situation would, on the other hand, consider that, even when confounding by indication is present, the resulting information can be useful for other purposes. For example, it has been postulated that the reported association between β -agonists and asthma mortality was confounded by disease severity.²⁵ Even in the presence of such a bias, however, the fact that β -agonists were positively correlated with asthma deaths can be very useful information for physicians, as the amount of drug use becomes a good indicator of prognosis.

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,²⁷ which showed that in the placebo group the risks of death in the five 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.,

there is a third variable, or group of variables, that plays 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 strategies to be controlled for. As we have seen, Tables 43.1 and 43.2 present two hypothetical examples, one with confounding (Table 43.1) and the other one with interaction (Table 43.2). 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 43.1 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 43.1 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 43.2 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, as 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, such as 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. Effect modification, thus, is often useful for generating new hypotheses and is generally presented as a finding that may be worth being further studied.

Through these considerations it appears that confounding is a nuisance and a threat in pharmacoepidemiology because it may affect the validity of the results. On the other hand, effect modification indicates a variation in the drug effect, according to different levels of a third variable. If it is confirmed, this change of effect may be the source of important findings regarding the use of the drug and its mechanism of action. Effect modification is thus an important piece of information that deserves high consideration, adequate study, and careful interpretation.

Effect Modification by Dose or Drug Potency

Different dosages and potencies 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 and biases the results toward the null. Spitzer *et al.*, for instance, concluded that use of both fenoterol and albuterol were associated with an excess risk of death in asthmatic patients, probably in relation to the severity of the patient's condition.²⁵ Table 43.3 shows that the dichotomous classification of drug exposure was associated with a threefold excess risk of death with fenoterol compared to albuterol. However, taking into account (1) the number of inhalers used, and also (2) 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 appear similar with the two drugs. This demonstrates the importance in pharmacoepide-

Table 43.3. 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 vs not exposed	fenoterol	9.1	3.0–0.1
	albuterol	3.8	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	113.2	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

* From reference 23, with permission

miology 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, which all result in selecting a study population that accurately represents the target population concerning the drug exposure–outcome relationship.

- Random sampling of the cases and controls (or exposed and nonexposed patients) to be included in the study from the source population.
- Systematically recruiting a series of consecutive patients (to prevent self-selection).
- Adopting a well codified accrual procedure (to be adapted according to the nature and severity of the disease). Having a geographic definition

of the incident cases goes a long way toward reducing 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 establish the reason and, if possible, to measure their health status.
- Selecting only incident cases of the condition.
- Random allocation of drug exposure, which prevents both 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* is the most important strategy, as it is easier to be neutral when it is not known who is exposed, who is sick, and what are the objectives of the study. In a cohort study, the data collector should be blind as 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 blind 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. Blinded assessment of the information is often difficult, however, in nonexperimental designs.
- *Standardization of the measurement process* for both cases and controls, or exposed and unexposed people, is an essential step when implementing a pharmacoepidemiology 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 with 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 have been developed to deal with confounding when working with large although incomplete (no information on confounding) databases, a frequent condition 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.

Randomization

Random allocation of exposure should equalize the distribution of all potential confounders, even unknown ones, across the different levels of drug exposure. *Randomization* is thus aimed at making the two groups perfectly similar apart from the independent intervention variable under assessment as exposure. This shows that the problem of confounding is in this way similar to selection bias. The distinction between the two situations may be subtle and, as Rothman wrote, a practical distinction between confounding and other biases is to consider bias as confounding if it can be controlled in the data analysis.¹

Matching

Matching is another way to control for confounding at the design stage. The objective of matching for both cohort or 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 to match for; it may then become costly and time consuming. In case-control studies matching may also lead to a new distorting phenomenon called "overmatching:" similarity with regard to exposure may also be related to disease status. In this case, there is artifactually no apparent difference between the exposure level of cases and controls (with regard to the exposure of interest) because of the matching.

Restriction

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 gives one 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 that are, respectively, exposed (a_i) and not exposed (b_i) to the drug.

There are two methods of standardization: direct and indirect. Table 43.4 shows the principle of each method and the requirements for their

computation. The number of factors that can be managed by standardization is limited (i.e., two or three). Pharmacoepidemiology research usually requires the manipulation of more than three factors and often, therefore, requires the use of other techniques such as modeling.

Stratification

Stratification is another way of obtaining a summary rate ratio. 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.^{1,7-10} 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

Table 43.4. Standardizing because of difference in age distribution

Methods of standardization	Requirements	Principles
Direct	<p>age-specific rates in each age stratum¹</p> <p>a standard population with its age distribution²</p>	<p>The age-specific rates for the two compared groups are applied to the standard population. For each age stratum of the standard population an expected number of events is obtained for each observed group; it is the one that would have been observed if the standard population had the age-specific rates of the observed groups. The sum of these expected cases in each group is then divided by the size of the standard population. This provides one age-adjusted mortality rate for each group, which may then be compared.</p>
Indirect	<p>distribution by age in each of the two compared groups</p> <p>age-specific rates of a standard population</p>	<p>The population of each stratum of the two compared groups is multiplied by the age-specific rate of a standard population. The computation gives an expected number of events in each group, which is the number that would have occurred if these groups had the age-specific rates of the standard population. In each studied population the expected number is then divided by the observed number of events. The values are called standardized mortality (or incidence) ratios</p>

Mantel and Haenszel for the odds ratio in case–control studies,²⁸ 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. When this is the case, no or very little confounding remains in each stratum. 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.¹⁵

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 pre-specified 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. Modeling is a very powerful technique, which requires sophisticated skills in biostatistics and epidemiology. Most of

the important models are derived from the general linear model.

Dealing with Confounding When Working with Large Drug Databases

Standardization, stratification, and modeling all require an accurate measurement of the confounding variable. Recently, a large number of pharmacoepidemiology studies have been performed using large databases of previously collected data (see also Chapters 15–23). 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 and the risk of myocardial infarction. Databases are likely to provide accurate figures for oral contraceptive (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 prevents studying this association, because the results will obviously be biased: OC users are also heavy smokers. When information on the confounder is available for the cases only,²⁹ Ray showed how to use this information to assess the presence of confounding. If it can be assumed there is no effect modification then, if there is no association between the confounder and exposure among the cases, there is no confounding, and a valid analysis can be carried out. This suggestion is interesting because pharmacoepidemiologists often have more information for the cases (because of hospitalization records) than for the controls (see also Chapter 44).

When information on confounders is not available, it is possible to simulate the effect that the confounder could have by incorporating into a model the information related to the strength of the association of the confounding variable with both the outcome and the exposure of interest, as well as the proportion of people exposed to the confounder. Fine and Chen,³⁰ for instance, reconsidered the relationship between the risk of pertussis vaccine and the occurrence of sudden

infant death syndrome (SIDS). Most studies have shown a significant protective effect of the vaccine, with relative risks as low as 0.15 and a 95% confidence interval of 0.05 to 0.45.³¹ Despite the strength of this association, almost no causal inference has been drawn from these results; it is generally assumed that other external factors might be responsible for either avoidance or delay of vaccination, and further development of SIDS. Fine and Chen identified a list of seven potential confounders that had been shown to be related to "failure or delay in receiving vaccines," and also were risk factors for SIDS. They further focused on what they identified as "contraindications to vaccine" and built a model for correcting the crude relative risk. They could study the variation of the adjusted relative risk for different values of (i) the percentage of children with vaccination, (ii) the relative risk of SIDS associated with vaccination, and (iii) the percentage of vaccination among children with and without contraindications, respectively. These models are certainly interesting for assessing the magnitude of the problem; they are, however, limited for providing an accurate adjusted figure.

In 1982, White³² and Walker³³ suggested sampling a fraction of the study population to gather information about confounding variables, and using this information in the analysis to obtain covariable-adjusted estimates of the parameters of interest. This approach, referred to as "two stage sampling,"^{34,35} was further developed by Cain and Breslow³⁶ 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 over-representation 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 which must be corrected in the analysis. The two-stage design permits the detection of, and adjustment for, confounding. Interaction can also be evaluated. Recently, Schaubel *et al.* proposed software for sample size estimation for two-stage sampling.³⁷

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.

It is also interesting to consider bias and confounding in pharmacoepidemiology under a dynamic perspective. We have already described the change in referral patterns after the publication of a single case report. 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 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. This change is normal and expected, as obtaining a more accurate definition of the treatment target population is an objective of pharmacoepidemiology research.³⁸ Any further studies of the effect of the drug may then be biased by this new reason for prescribing. We may therefore view confounding by indication as a natural and positive conse-

quence of integrating the results of research into medicine, 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 more accurately (see also Chapter 39). 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 represents thus a highly promising domain of research with regard to the study of drug effects. The development of population pharmacokinetics and pharmacodynamics, as well as pharmacogenetics, should also provide useful information for interpreting pharmacoepidemiology results with regard to drug exposure.

REFERENCES

- Rothman KJ. *Modern Epidemiology*. Boston, MA: Little, Brown, 1986.
- Committee on Safety of Medicines and the joint Committee on Vaccination and Immunization. *Whooping Cough*. London: Her Majesty's Stationery Office, 1981.
- Pollock TM, Morris J. A 7-year survey of disorders attributed to vaccination in North West Thames region. *Lancet* 1983; **1**: 753–7.
- Shields WD, Nielsen C, Buch D, Jacobsen V, Christenson P, Zachau-Christenson B, Cherry JD. Relationship of pertussis immunization to the onset of neurological disorders. A retrospective epidemiologic study. *J Pediatr* 1988; **113**: 801–5.
- Walker AM, Jick H, Petra DR, Knauss TA, Thompson RS. Neurologic events following diphtheria–tetanus–pertussis immunization. *Pediatrics* 1988; **81**: 345–9.
- Griffin MR, Ray WA, Morimer EA, Fenchel GM, Schafner W. Risk of seizures and encephalopathy after immunization with the diphtheria–tetanus–pertussis vaccine. *J Am Med Assoc* 1990; **263**: 1641–5.
- Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research*. New York: Van Nostrand Reinhold, 1982.
- Feinstein AR. *Clinical epidemiology: The architecture of clinical research*. Philadelphia, PA: Saunders, 1985.
- Kramer MS. *Clinical epidemiology and biostatistics*. Heidelberg: Springer, 1988.
- Checkoway H, Pearce NE, Crawford-Brown DJ. *Research methods in occupational epidemiology*. New York: Oxford University Press, 1989.
- Walker AM. Observation and inference. *An introduction to the methods of epidemiology*. Chestnut Hill, MA: Epidemiology Resources, 1991.
- Miettinen OS. Confounding and effect modification. *Am J Epidemiol* 1984; **100**: 350–3.
- Copeland KT, Checkoway H, McMichael AJ, et al. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977; **105**: 488–95.
- Miettinen OS, Cook F. Confounding: essence and detection. *Am J Epidemiol* 1981; **114**: 593–603.
- Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980; **112**: 564–9.
- Greenland S, Robins JM. Confounding and misclassification of covariates. *Am J Epidemiol* 1985; **122**: 495–506.
- Van der Kroef C. Reactions to triazolam. *Lancet* 1979; **2**: 526.
- Schoght B. Paranoid symptoms associated with triazolam. *Can J Psychiatry* 1985; **30**: 462–3.
- Weiburg JB, Sachs F, Falk WE. Triazolam-induced brief episodes of secondary mania in a depressed patient. *J Clin Psychiatry* 1987; **48**: 492–3.
- Sunter JP, Bal TS, Cowan WK. Three cases of triazolam poisoning. *Br Med J* 1988; **297**: 719.
- The International Agranulocytosis and Aplastic Anemia Study. Risks of agranulocytosis and aplastic anemia: a first report of their relation to drug use with special reference to analgesics. *J Am Med Assoc* 1986; **256**: 1749–57.
- 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.
- Strom BL, Carson JL, Morse ML, et al. The effect of indication on hypersensitivity reactions associated with zomepirac sodium and other nonsteroidal antiinflammatory drugs. *Arthritis Rheum* 1987; **30**: 1142–9.
- Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983; **2**: 267–71.
- Spitzer WO, Suissa S, Ernst P, et al. The use of beta-agonists and the risk of death and near death from asthma. *New Engl J Med* 1992; **326**: 501–6.
- Petri H, Urquhart J, Herings R, et al. Characteristics of patients prescribed three different inhalational beta-2 agonists: an example of the channeling phenomenon. *Post-Mark Surveil* 1991; **5**: 57–65.
- Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *New Engl J Med* 1980; **303**: 1038–41.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **22**: 719–48.

29. Ray W, Griffin MR. Use of Medicaid data for pharmacology. *Am J Epidemiol* 1989; **129**: 837–49.
30. Fine PEM, Chen RT. Confounding in studies of adverse reactions to vaccines. *Am J Epidemiol* 1992; **136**: 121–35.
31. Walker AM, Jick H, Perera DR, *et al.* Diphtheria–tetanus–pertussis immunization and sudden infant death syndrome. *Am J Public Health* 1987; **77**: 945–51.
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.
33. Walker AM. Anamorphic analysis: sampling and estimation for covariate effects of both exposure and disease are known. *Biometrics* 1982; **38**: 1025–32.
34. Zhao LP, Lipsitz S. Design and analysis of two-stage studies. *Stat Med* 1992; **11**: 769–82.
35. Wacholder S, Carroll RJ, Pee D, Gail MH. The partial questionnaire design for case–control studies. *Stat Med* 1994; **13**: 623–4.
36. Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. *Am J Epidemiol* 1988; **28**: 1198–206.
37. 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.
38. Collet JP, Boissel JP. Populations: evaluation and treatment. The need for a better definition. *Eur J Clin Pharmacol* 1991; **41**: 267–71.

44

Novel Approaches to Pharmacoepidemiology Study Design and Statistical Analysis

SAMY SUISSA

Department of Epidemiology and Biostatistics, Department of Medicine, and Pharmacoepidemiology Research Unit,
McGill University, Montreal, Canada

INTRODUCTION

The last 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 precise applications, 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. Even if all the data are available, the fact that exposure, namely drug use, varies over time, and often involves multiple uses, implies very complex measures of exposure and formidable technical challenges in data analysis. Consequently, it becomes indispensable to use strategies based on the collection and analysis of data for a sample of the cohort. The nested case-control and case-cohort techniques are discussed for both internal and external comparisons of adverse event rates. We also discuss the role of time in sampling within the cohort.

Second, we describe techniques of design and analysis for situations where only partial data are

available on confounders. The first, an approach developed around specific problems in pharmacoepidemiology, is a practical method that permits the investigator to assess whether a factor is a confounder, and to adjust for this confounder, when data on this factor are only available in the cases and not in the controls of a case-control study. This situation is commonly encountered in pharmacoepidemiology, particularly when using computerized databases where data, in both quantity and quality, are more readily available for the cases than noncases. We also briefly introduce the two-stage sampling technique, which deals with confounder data available on a sample of the cases and controls of the study population.

Third, we describe new designs that use within-subject comparisons to estimate risk, namely the case-crossover and case-time-control designs. These designs, dealing with the study of transient drug effects on the risk of acute adverse events, are useful when there is uncertainty about the proper selection of controls under the traditional case-control approach or when unmeasured confounders are deemed to be important. 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 channeling of drugs.

CLINICAL PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMOLOGY 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. Whereas the more common conditions are usually studied prior to drug marketing using experimental research designs such as randomized clinical trials, for the study of uncommon adverse conditions, the

mainstay of pharmacoepidemiology, we must rely on methods based on nonexperimental designs. The greatest challenge of this field is then to quantify the risk of a drug accurately, relative to one or a variety of alternatives. Three 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,¹ while the Nurses' Health Study cohort used 121 700 subjects to study the effect of oral contraceptives on the risk of cardiovascular diseases.² As described in Part III of this book, hurdles in forming these massive cohorts have been reduced by the use of computerized databases. Such databases have in fact revolutionized risk assessment research in pharmacoepidemiology, where conclusive information about a drug's potential risks cannot be delayed by the lengthy process of classical epidemiologic methods. Nevertheless, even if many of these data are already computerized, they remain expensive and time-consuming to collect in such mega-sized cohorts. Moreover, drug exposure often varies over time and involves multiple agents, which complicates the analysis of cohort data. Accordingly, efficient designs to sample a manageable number of study subjects within such cohorts have been devised and can be used effectively in pharmacoepidemiology, providing accurate results more rapidly and at less expense. The first part of this chapter deals with these more efficient designs.

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.³ A few, however, still suggest that this is still an unresolved issue, as the association may be confounded by the mother's history of spontaneous abortions and of bleeding during pregnancy.⁴ Analysis of such crucial studies requires a thorough knowledge and accurate measurements of the potential clinical confounders. Since data on confounders may be difficult to obtain on all subjects, a novel approach based on confounder data measured only on cases has been proposed and applied successfully in several studies. Moreover, a formula to estimate the corresponding adjusted rate ratios has been devised. We describe this technique here, since it is particularly suited to pharmacoepidemiology, where this situation is most often met. In addition, the two-stage sampling technique, which measures confounders on a sample of the case-control study population, will also be introduced as an alternative tool to address this challenge.

Third, pharmacoepidemiology is frequently faced with the assessment of the risk of rare acute adverse events resulting from transient drug effects. For example, we may wish to study the risk of ventricular tachycardia in association with the use of inhaled β -agonists in asthma. This possible effect has been hypothesized on the basis of clinical study observations of hypokalemia and prolonged Q-T intervals in patients after β -agonist exposure.⁵ These unusual cardiac deviations were observed only in the 4 hour period following drug absorption. Although the case-control approach can be used to address this question, the acuteness of the adverse event and the length of the drug's effect, as well as difficulties in determining the timing of drug exposure, induce uncertainty about the proper selection of controls. Moreover, confounding by indication may often be a problematic issue in such a design. We will review a recently proposed approach that addresses these difficulties, the case-crossover design. We also review the case-time-control design, which was devised to counter the time-trend biases inherent in the case-crossover

design. 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. These techniques, useful to describe drug use patterns and assess risk, are briefly discussed.

METHODOLOGIC PROBLEMS TO BE ADDRESSED BY PHARMACOEPIDEMIOLOGY 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). For example, to quantify the risk of an adverse event occurring at the rate of 2 per 10 000 per year with a precision of ± 1 per 10 000 with 95% probability necessitates a cohort of close to 80 000 subjects followed for 1 year (or 160 000 subjects followed for 6 months). This type of requirement explains the infrequent use of the cohort design in pharmacoepidemiology. Examples of authentic cohorts are scant. The Cancer Prevention Study II cohort of 1.2 million persons enrolled in 1982 was used to assess the effect of aspirin use on the risk of colon cancer.¹ The Nurses' Health Study cohort of 121 700 female nurses established in 1976 evaluated several associations, notably the effect of oral contraceptives on the risk of cardiovascular diseases.² The McGill Hodgkin's Disease cohort of 10 472 patients was formed to assess the effects of chemotherapy on the risk of second cancers⁶ and of coronary artery disease.⁷ The Oxford Family Planning Association cohort of 17 032 women assessed mortality associated with oral

contraceptive use.⁸ Other examples exist, but the list is short.

Complementing such authentic cohorts formed from actually following a group of “live” subjects over time, the “computerized” cohort has become a popular alternative. Most of Part III of this book is devoted to the description of national and regional computerized health databases, frequently used to form cohorts for drug risk assessment, from which we derive the distinguishing designation “computerized” cohort. In fact, because of the urgent nature of several drug risk assessment situations, the “live” cohort approach, with its prospective nature, cannot address the problem sufficiently rapidly in a majority of instances, unless of course if it happens to have been already formed and followed up for another purpose. Accordingly, the “computerized” cohort has become an indispensable instrument in the armamentarium of many investigators in pharmacoepidemiology.

In any cohort study, the cost, time, and resources necessary to collect data on all cohort members can be prohibitive. Moreover, even with computerized cohorts, most studies will need to supplement and validate data obtained from the computer databases with data from hospital records, medical records, and physician or patient interview questionnaires. When, as in the majority of instances, the cohort size is considerable, such additional data gathering can become a formidable task, if not next to impossible. More importantly, however, is when the exposure varies over time and includes several drugs. The analysis of such a cohort with time-dependent exposure measures can be infeasible, if not impossible. To counter these obvious cost, time, and feasibility constraints, designs based on sampling subjects within a cohort have been proposed and recently applied successfully in pharmacoepidemiology. These designs are based on the selection of all cases with an adverse event in the cohort, but differ in the selection of a small subset of noncases. Generally, they permit the precise estimation of relative risk measures with negligible losses in precision. Below, we discuss structural aspects of cohorts and present two sampling designs within a cohort, the nested case–control and case–cohort designs.

PARTIAL CONFOUNDER DATA

The most important limitation of the nonexperimental research methods frequently used in pharmacoepidemiology is the reservation about error in the final risk estimate. Invariably, questions are posed as to whether an important confounding factor which was not considered in the study design is biasing the reported results. A factor is considered a *confounder* if it is associated with the adverse event irrespective of exposure to the drug under study, and with exposure to the drug itself. Chapter 43 addresses this issue.

Essentially, two classical solutions exist to address the problem of confounding variables in nonexperimental studies, once all such potential confounders have been identified *a priori*. We will use the case–control design to illustrate the problem, although the principles discussed in this section apply equally well to a cohort design. The first solution is to design the study to account for the confounder, e.g., selecting controls that are matched to the cases with respect to all these confounding factors and then using the appropriate corresponding techniques of analysis for matched data. The second solution is to select controls unmatched with respect to these confounding factors, but to measure these confounders in the course of the study for all subjects and use statistical techniques, based on either stratification or regression, permitting one to remove their effect on the risk from the effect of the drug under study. The advantage of the first approach over the second is that, for strong confounders, fewer subjects will be needed to attain a desired level of power and simpler techniques of data analysis will be necessary. The second approach will permit one to assess the relative contribution of the confounders to the risk, which is impossible with the first approach. Irrespective, both standard approaches require the measurement of each confounding factor for each subject in the study, case and control, whether at the design stage for the first or at the analysis phase for the second solution.

However, it is at times impossible to obtain data on certain important confounding variables. This limitation could be fatal to a study that is already

based on fragile nonexperimental methodology. A frequent situation encountered in pharmacoepidemiology 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 “computerized” 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, as is usually the practice with “computerized” studies, it is unlikely they were hospitalized and, even if they were, probably not as extensively as the cases. The controls will therefore not be able to provide comparable data on confounders on the basis of medical charts only. Since, in such studies, we rarely contact subjects directly but rather rely exclusively on charts and records to supplement the computerized information, confounder data will typically only be available in the cases, and not in the controls. This chapter will describe a strategy of analysis to assess whether a factor is a confounder based on data available uniquely for the cases and the formula to estimate the rate ratio adjusted for this partially available confounder. Another closely related situation is when the confounders can be measured on both cases and controls, but resources only allow this measure for a sample of subjects. The two-stage sampling approach has been advanced as a strategy to efficiently select this sample and obtain adjusted estimates of the rate ratio. This approach will also be briefly described.

WITHIN-SUBJECT DESIGNS

When conducting a conventional case-control study (as distinct from the one nested within a cohort), the selection of controls is usually the most challenging task. The fundamental principle used in this process is that selected controls should be representative of the source population which gave rise to the cases,⁹ a principle which derives from the case-base paradigm.¹⁰ Although elementary in theory, the principle is often extremely complex to implement in practice, irrespective of

whether one decides to select population or hospital controls.

For population controls, specific hurdles make it difficult to apply the principle. First, we can often expect significant nonresponse or nonacceptance rates in the control group, particularly if the data collection instrument is burdensome, as it frequently is. This is rarely acceptable since the reasons for acceptance of participation as a control are not easily known and could be associated with exposure to the drug of interest, while the corresponding case selection is usually comprehensive with essentially complete response and acceptance rates. Consequently, the source population could have been well identified on theoretical grounds, but the practical results of the sampling exercise could result in a control population unrepresentative of this source population.

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, which was conjectured after observations of hypokalemia and prolonged Q-T intervals in patients within the 4 hour period following β -agonist absorption,⁵ a case-control study may be attempted. Indeed, one would first select cases with this adverse event and easily probe whether they took the drug during the 4 hour 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 hours.” 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. Choosing an array of diagnoses suitable for control subjects is not simple. Acute conditions could be associated with an elevated prevalence of use of the drug under study. Alternatively, hospitalizations for chronic diseases could be planned with specific contraindications against use of the drug. In addition, problems related to timing as discussed above can be as or more complicated in this context.

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. Moreover, other within-subject techniques based solely on prescription drug databases will also be discussed.

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.¹¹ For illustrative purposes, consider the subcohort of newly diagnosed diabetics only. The eligibility criteria may be one or more of disease status (insulin-dependent diabetes mellitus), age (diabetics less than 40 years of age), drug use (regular users of

insulin), geographical location (urban/rural residence), etc. Entry into the cohort may be defined by calendar time (spanned by the study, e.g., any time after 1 January 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 calendar time (e.g., 31 December 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 44.1. 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 1 January 1978), depicting the chronological nature of the

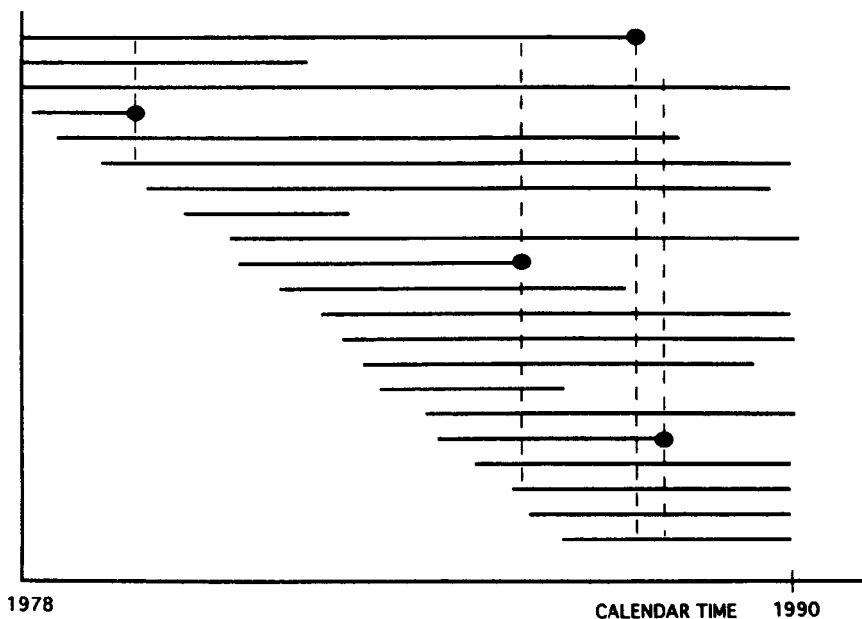


Figure 44.1. Illustration of a variable-entry cohort of 21 subjects followed from 1978 to 1990 with four cases (●) occurring and related risk-sets (---).

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 44.2 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 as to 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 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 44.1 and 44.2 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 44.1, for example, case 1 has in its risk set only the first six subjects to enter the cohort, while in Figure 44.2, 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

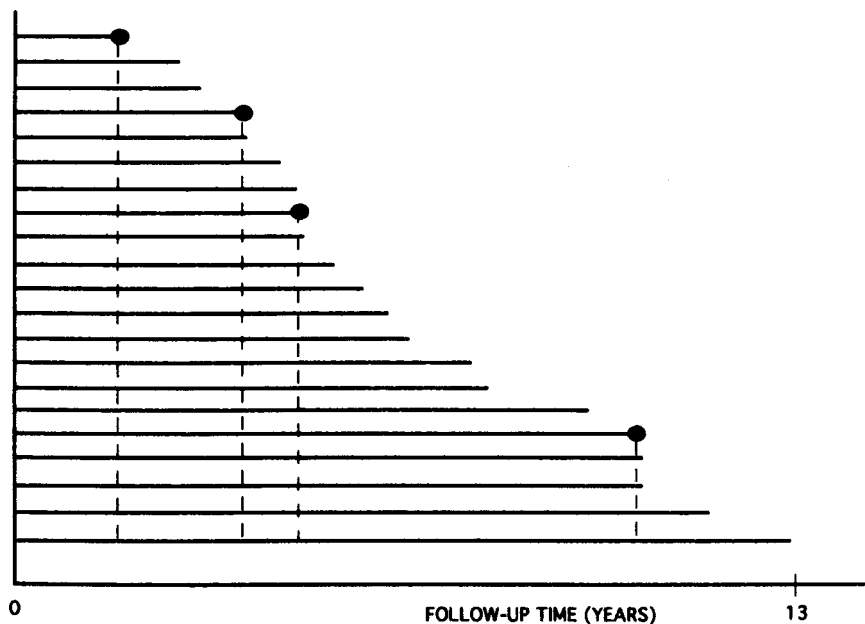


Figure 44.2. Illustration of fixed-entry cohort representation of the cohort in Figure 44.1, with new risk-sets (---) for the four cases.

for the formation of risk sets and the second form for data analysis. 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, duration of disease for another.

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 their own zero time defined by entry date, analyzing duration within each subcohort. We could, for example, partition the cohort displayed in Figures 44.1 and 44.2 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 44.1 and 44.2. 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,¹² who proposed an unmatched selection of controls and called it a synthetic retrospective study. It was developed further and formalized by Liddell *et al.*,¹³ in the context of a cohort study of asbestos exposure and the risks of lung cancer and mortality. The nested case–control design involves three steps:

1. defining the cohort’s time axis;
2. selecting all cases in the cohort, i.e., all subjects with an adverse event;
3. forming all risk sets corresponding to the cases; and
4. randomly selecting one or more controls from each risk set.

Figure 44.3 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 noncases 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.¹⁴ 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. We faced this problem in a study of the risks of severe adverse events in asthma associated with the use of inhaled β -agonists.¹⁵ 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 feasible collection of additional data from hospital charts and questionnaires sent to all physicians who saw these patients, it was necessary to sample from the cohort. These additional data were specifically focusing on the two-year period prior to the risk set date. A standard nested case–control sample of six controls per case, as described above, would have produced several subjects with more than one occurrence in the sample. This would have been problematic vis-à-vis the questioned physicians, for example, who would have had to respond to questions about the same patient’s asthma severity in different two-year periods, a clearly confusing and potentially unreliable data collection scheme. To circumvent this difficulty, we stratified the cohort according to various potential confounding factors, namely age, area of residence, social assistance, prior asthma hospitalization, and, of course, calendar date of entry into the cohort. This extremely fine stratification resulted in 129 subcohorts, one for each case. We thus selected between two and eight controls per case (some

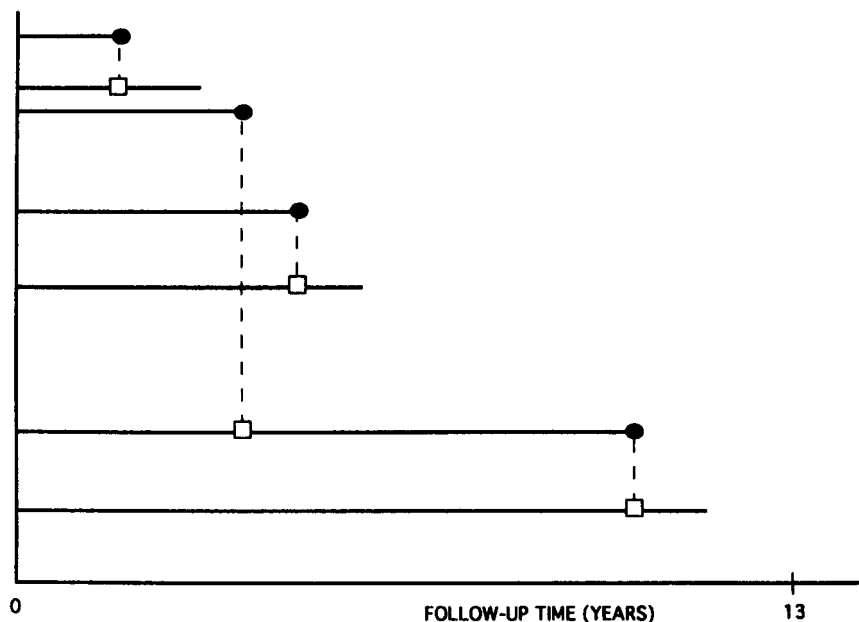


Figure 44.3. Nested case-control sample of one control (□) per case (●) from the cohort in Figure 44.2.

risk sets contained only two eligible controls, mostly because of the matching by prior hospitalization). 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. This stratified nested case-control design strategy is commonly needed in pharmacoepidemiology since calendar time matching for cohort entry is often essential to account for disease duration and drug exposure time trends.

The analysis of data from a nested case-control study must conserve the matched nature of the selection of cases and controls, particularly if the risk of the adverse event changes with disease duration and drug exposure varies in calendar time. This also applies to the stratified nested case-control design, which is based on other matching criteria as well. The method of analysis is identical to that of a conventional matched case-control study, not nested within a cohort. The conditional logistic regression method for this design is appropriate, as it uses the risk set as the fundamental unit of analysis, in agreement with the proportional hazards model of the full

cohort.¹⁶ Simple formulae exist to estimate the relative risk for 1:1 matching.⁹ When more than one control is matched to each case, however, sophisticated computer packages such as EGRET¹⁷ are required to fit the necessary conditional logistic regression model. This technique was used in our nested case-control study of β -agonist risks in asthma,¹⁵ where the number of controls per case varied between two and eight.

The question of the required number of controls per case is important. Although selecting one control per case will greatly simplify the data analysis, a large number of cases will be required to attain an acceptable level of power. Since the number of cases in the cohort is fixed and cannot be increased to satisfy this requirement, the only remaining alternative is to increase the control-to-case ratio. Tables are given by Breslow and Day.¹⁸ It can be readily seen from these sample size tables that the gain in power is significant for every additional control up to four controls per case, but becomes negligible beyond this ratio. For example, if we consider an exposure prevalence in the controls to 30% and target a relative risk of 2 with 5% significance and 80% power, the required

numbers of cases are 122, 90, 74, 65, and 62 respectively, for 1:1, 2:1, 4:1, 10:1, and 20:1 control-to-case ratios. These translate to total study sizes (cases and controls) of 244, 270, 370, 715, and 1302, with clear cost implications and related optimality decisions. Of course, the number of cases in a cohort is frequently fixed *a priori* by the study constraints, thus eliminating this option to increase the number of cases. However, although this general rule of an optimal 4:1 control-to-case ratio is appropriate in the majority of instances, one should be prudent when exposure to the drug under study is infrequent, when the hypothesized relative risk moves further from unity, or when several factors or other drugs are being assessed simultaneously. In these situations, the ratio could easily be required to increase to ten or more controls per case. This is the case in two recent studies. In our study of the risks of fatal or near-fatal asthma associated with the use of inhaled β -agonists,¹⁵ the low rate of use of fenoterol, believed to be around 5%, dictated the selection of up to eight controls per case. In a proposed study of the cardiovascular risks of oral contraceptives,¹⁹ the relative infrequency of newer oral contraceptives with lower estrogen dose and new progestins, coupled with the strong confounding effect of age, led to a requirement of ten 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 contrast exposure to a drug to no exposure, or to some average exposure. This is not possible using methods for internal comparisons when all subjects in the cohort are exposed to the drug under study. Instead, external comparisons are performed, 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) when the adverse event is death, or the standardized incidence rate (SIR) when it is not. Techniques to estimate these measures using the full cohort are described in most textbooks of epidemiology.^{9,10}

The nested case-control design, however, is not a simple random sample from the cohort and thus cannot use the same techniques to estimate these measures. Indeed, it is evident from Figure 44.3 that those cohort members with the longest follow-up have a greater chance of being selected in the nested case-control sample, since they belong to all the risk sets. If their drug exposure pattern is different from that of other cohort members, the members of the nested case-control sample will not be representative of the cohort, which may substantially bias the resulting analyses. The appropriate method to perform external comparisons using data from a nested case-control design has been described.²⁰ It uses knowledge about the sampling structure to yield an unbiased estimate of the adverse event rate in the full cohort, thus permitting the estimation of the necessary standardized relative measure with respect to the selected external population.

The Case-Cohort Design

The first recognized application of a sampling design we currently call *case-cohort* was made by Hutchison,²¹ in performing external comparisons of leukemia rates in patients treated by radiation for cervical cancer. It was ultimately developed and formalized by Prentice,²² who coined the name *case-cohort*. Although recent, this design has already been used effectively in some drug risk studies.²³⁻²⁶ The case-cohort design involves two steps: (1) selecting all cases in the cohort, i.e., all subjects with an adverse event; and (2) randomly selecting a sample of predetermined size of subjects from the cohort. Figure 44.4 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 to either the established cohort or case-control methods, the techniques are in

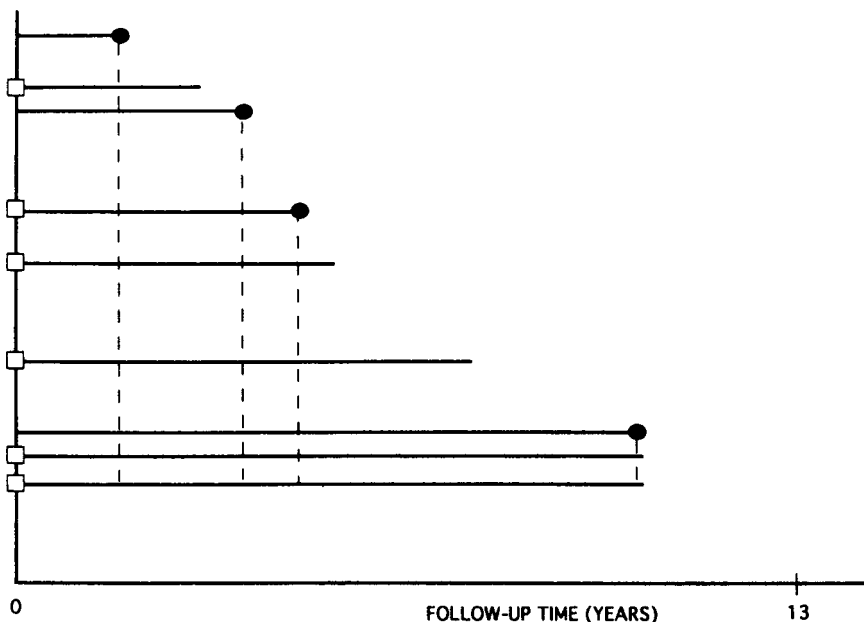


Figure 44.4. Case-cohort sample with six controls (□) from the cohort in Figure 44.2.

fact distinct, each requiring specific software. This is because the method of analysis for a case-cohort sample is extremely complex, as it must take into account the overlap of cohort members between successive risk sets induced by this sampling strategy. This handicap has severely limited the use of the case-cohort design. However, a statistical software package called EPI-CURE²⁷ 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.²⁸

The first advantage of the case-cohort design is its capacity to use the same sample to study several different events. Indeed, the cases can be split into several subcategories and each can be analyzed with the same “control” subcohort. In contrast, the nested case-control design requires different control groups for each type of event because the selection depends on event times. For example, the

β -agonist risk nested case-control study had two distinct control groups, one of size 233 for the 44 asthma deaths, the other of size 422 for the 85 asthma near-deaths.¹⁵ Another useful advantage is that the case-cohort design permits one to change the primary time axis of analysis from calendar to disease time and vice versa, depending on either the assumed model or the targeted outcome. This is not possible with the nested case-control study, where the primary time axis must be set *a priori* to permit the risk-set construction. This is less of a problem in pharmacoepidemiology, however, where the cohort can be divided into subcohorts of successive calendar time, as was discussed earlier. Yet another example is its simplicity in sampling, which has advantages in both comprehensibility and computer programming. Finally, external comparisons are simple to perform with the case-cohort approach.²⁹

The nested case-control design does have some advantages. The first is the simplicity of power calculation, or equivalently sample size determination. The nested case-control design is independent of the size of the cohort, while for the case-cohort design knowledge about overlap in risk sets

is essential, thus greatly complicating these calculations. Second, data on time-dependent exposure and covariates need only be collected up to the time of the risk set for the nested case-control study, while the collection must be exhaustive for the case-cohort. Finally, despite the accessibility of software for data analysis of case-cohort data, these can quickly become surpassed and even infeasible with larger sample sizes and time-dependent exposures. In this situation, the nested case-control design, with its single risk set per case, is not only advantageous but also the only solution. A recent 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.³⁰

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 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 Griffin³¹ and was used in the context of a study of nonsteroidal anti-inflammatory drugs and the risk of fatal peptic ulcer disease.³² The strategy is based on the definition¹⁶ 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.

As an example of this approach, we use data from a case-control study of estrogen use and the risk of endometrial cancer.³³ These data, restricted to the effect of former estrogen exposure, were used previously to illustrate confounding³⁴ by simulating hypothetical stratifications by a third factor. They are displayed in Figure 44.5. The crude odds ratio between estrogen use and endometrial cancer is 3.3 in this study of 17 cases and 60 controls. When the study population is partitioned according to the hypothetical potential confounding factor parity, the odds ratios are 1.1 in the low parity group and 1.0 in the high parity group, indicating a strong confounding effect of parity. This is a result of the strong association, using the odds ratio as the measure of association, between parity and estrogen exposure in the control group (OR = $8 \times 39 / (3 \times 10) = 10.4$) and the strong association between parity and endometrial cancer in subjects exposed to estrogen (OR = 11.3) and subjects not exposed (OR = 9.8). The bottom section of the figure displays the data *for the cases only*, where the odds ratio between estrogen exposure and parity is 12, demonstrating analogously the strong confounding role of parity. If the parity data were only available in the cases, this finding would invalidate the crude odds ratio of 3.3. Conversely, the same example could be used with smoking as another hypothetical confounder, as displayed in Figure 44.6. It is clear that the crude odds ratio of 3.3 is not affected by stratification by smoking, which is therefore not a confounder. This is confirmed by the lack of association between estrogen exposure and smoking *in the cases only*, with an odds ratio of 1.1.

	Case	Control		
User	10	18		
Nonuser	7	42		
	OR = 3.3			
	┌───────────┴───────────┐			
	┌───────────┐		┌───────────┐	
	Low parity		High parity	
	Case	Control	Case	Control
User	9	8	1	10
Nonuser	3	3	4	39
	OR = 1.1		OR = 1.0	

In cases only:

	Low	High
User	9	1
Nonuser	3	4
	OR = 12	

Figure 44.5. Data from case-control study of former estrogen use and endometrial cancer with hypothetical stratification by parity (confounding present).³³

	Case	Control		
User	10	18		
Nonuser	7	42		
	OR = 3.3			
	┌───────────┴───────────┐			
	┌───────────┐		┌───────────┐	
	Low parity		High parity	
	Case	Control	Case	Control
User	6	10	4	8
Nonuser	4	22	3	20
	OR = 3.3		OR = 3.3	

In cases only:

	Low	High
User	6	4
Nonuser	4	3
	OR = 1.1	

Figure 44.6. Data from case-control study of former estrogen use and endometrial cancer with another hypothetical stratification by parity (no confounding).³³

This strategy to assess confounding is extremely valuable for several case-control studies in pharmacoepidemiology since, if confounding is excluded by this technique, crude methods of analysis can be used to obtain a valid estimate of the odds ratio. However, 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.³⁵ The adjusted odds ratio is given by

$$OR_{adj} = P_0(w - y) / [(1 - P_0)y]$$

where $y = \{v - [v^2 - 4(r - 1)rx]^{1/2}\} / [2(r - 1)]$, $v = 1 + (r - 1)(w + x)$ 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.³⁵

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.³⁶ In this study, the 30 cases provided data on theophylline use, as well as on smoking, possibly an important confounder. On the other hand, the 4080 controls only had data available on theophylline use and not on smoking. Table 44.1 displays the data from this study. The crude odds ratio between theophylline use and

Table 44.1. Data from a case-control study of theophylline use and cardiac death in asthma, with the smoking confounder data missing for controls³⁶

	Cases		Controls	
	Use	Non use	Use	Non use
All subjects	17	13	956	3124
Stratified by smoking:				
smokers	14	5	*	*
non-smokers	3	8	*	*

*These frequencies are missing for controls.

cardiac death is 4.3 $((17/13)/(956/3124))$. Because of the missing data on smoking, it is only possible to partition the cases, but not the controls, according to smoking. The odds ratio between theophylline use and smoking among the cases is estimable and found to be 7.5 $((14/5)/(3/8))$, thus indicating that smoking is indeed a strong confounder. An external estimate of smoking prevalence among asthmatics, obtained from a Canadian general population health survey, is 24%. Using this estimate, the adjusted odds ratio is 2.4, much lower than the crude estimate of 4.3, with 95% confidence limits 1.0–5.8.

An alternative approach is available when the confounders can be measured on both cases and controls, albeit only for a sample of subjects. This technique, developed more than a decade ago but not widely used, is the two-stage sampling approach, in which stage 1 is 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 sample. This situation is common with database studies, where the database provides data on exposure and outcome for all subjects, but confounders are missing and need to be obtained directly on a subset of subjects. The balanced design, wherein an equal number of individuals is selected from each exposure/outcome category, is usually the most efficient strategy by which to select the stage 2 sample. This method has recently been used in pharmacoepidemiology.³⁷ 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. In this instance, validating the case status of all cases would be inefficient when most cases are unexposed. Validating all exposed cases and a sample of the unexposed cases turns out to be an efficient means of estimating the rate ratio.³⁸

WITHIN-SUBJECT DESIGNS

When dealing with the study of transient drug effects on the risk of acute adverse events, Maclure³⁹ 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. This is a design where

comparison 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

To carry out a case-crossover study, three critical points must be considered. First, the study must necessarily be dealing with an acute adverse event that is alleged to be the result of a transient drug effect. Thus, drugs with regular patterns of use which vary only minimally between and within individuals are not easily amenable to this design. Nor are latent adverse events, which only occur long after exposure. Second, since a transient effect is under study, the effect period (or time window of effect) must be precisely determined. For example, in our hypothetical study of the possible acute cardiotoxicity of inhaled β -agonists in asthmatics (see above), we identified this effect period to be 4 hours after having taken the usual dose of two inhalations of 100 mcg of the product. An incorrect specification of this time window can have important repercussions on the risk estimate, as we will show below in the example. Third, one must be able to obtain reliable data on the usual pattern of drug exposure for each case, over a sufficiently long period of time. For our example, we could seek the frequency of use of β -agonists during the year preceding the adverse event.

The case-crossover study is simply a crossover study *in the cases only*. The subjects alternate at varying frequencies between exposure and non-exposure to the drug of interest, until the adverse event occurs, which happens for all subjects in the study, since all are cases by definition. With respect to the timing of the adverse event, each case is investigated to determine whether exposure occurred within the predetermined effect period, namely within the previous 4 hours in our example. This occurrence is then classified as having arisen either under drug exposure or nonexposure on the basis of the effect period. Thus for each case, we have either an exposed or unexposed status which represents for data

analysis the first column of a 2×2 table, one for each case. Since each case will be matched to itself for comparison, the analysis is matched and thus we must create separate 2×2 tables for each case.

With respect to control information, the data on the average drug use pattern are necessary to determine the typical probability of exposure to the time window of effect. This is done by obtaining data for a sufficiently stable period of time. In our example, we may find out the average number of times a day each case has been using β -agonists (two inhalations of 100 mcg each) in the past year. Note that there are six 4 hour periods (the duration of the effect period) in a day. Such data will determine the proportion of time that each asthmatic is usually spending time in the effect period and thus potentially “at risk” of ventricular tachycardia. This proportion is then used to obtain the number of cases expected on the basis of time spent in these “at risk” periods, for comparison with the number of cases observed during such periods. This is done by forming a 2×2 table for each case, with the corresponding control data as defined above, and combining the tables using the Mantel–Haenszel technique as described in detail by Maclure.³⁹

We generated data for a hypothetical case-crossover study of ten asthmatics who experienced ventricular tachycardia. These were all queried regarding their use of two puffs of inhaled β -agonist in the last 4 hours and on average over the past year. The data are displayed in Table 44.2. The fact of drug use within the effect period for the event classification is straightforward. The usual frequency of drug use per year is converted to a ratio of the number of “at risk” periods to the number of “no risk” periods, the total number of 4 hour periods being 2190 in 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 hour period, is (0, 1, 365, 1825), while for the second case, who is exposed, it is (1, 0, 6, 2184). Using the Mantel–Haenszel technique to combine the ten 2×2 tables, the estimate of relative risk is 3.0 (95% CI: 1.2–7.6).

This method is sensitive to the specification of the time window of effect. For example, if this

effect period is in fact only 2 hours, then the data of Table 44.2 would be affected in two ways: some cases may no longer be considered exposed, and the exposure probabilities will change. By considering as unexposed cases 2 and 4, for instance, who may have been exposed 3 hours before ventricular tachycardia, and recalculating the appropriate exposure probabilities, the relative risk becomes 2.0 (95% CI: 0.3–12.0). On the other hand, if this effect period is in fact 6 hours long, then the data of Table 44.2 would be affected in two ways: some new cases may now be considered exposed, and the exposure probabilities will change. By considering as exposed cases 3 and 5, for instance, who may have been exposed 5 hours before ventricular tachycardia, and recalculating the appropriate exposure probabilities, the relative risk becomes 5.0 (95% CI: 2.0–12.2). The difference in the magnitude of the risk and the corresponding statistical significance between the various scenarios is indicative of the importance of the need for an accurate specification of the length of the effect period.

This method is extremely valuable when studying an acute adverse event that is alleged to be the result of a transient drug effect. Consequently, it excludes drugs with regular patterns of use, which vary minimally between and within individuals, or adverse events that can only result from long extended exposure. Moreover, the case-crossover design requires precise knowledge

about the effect period (or time window of effect), although the latter can be varied to investigate the optimum window to use. The design is also very useful when the selection of controls in the usual sense is uncertain. A significant advantage of this design is that it eliminates the problem of confounding by factors that do not change over time. It cannot, however, easily address the problem of confounding by factors that do change over time. In this instance, time-dependent data will be required for such confounders, a possibly difficult task. The case-crossover design is automatically free of selection bias, which occurs when controls are not representative of the base population from which the cases arose. However, although such control selection bias (in the usual control sense) is eliminated, case selection bias could be present if case selection is related to the exposure under study. 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. This source of bias can, however, be dismissed if one uses, for example, drug exposure data from computerized databases. Greenland⁴⁰ presented examples where the odds-ratio estimates from this approach can be biased. Nevertheless, this approach is being used successfully in several studies.^{41–44} It has also been adapted for application to the risk assessment of vaccines.⁴⁵

Table 44.2. Hypothetical data for a case-crossover study of β -agonist exposure in last 4 hours and the risk of ventricular tachycardia in asthma

Case No.	β -agonist use* in last 4 hours (a_i)	Usual β -agonist use in last year	Periods of risk (N_{1i})	Period 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

* Inhalations of 200 mcg: 1 = yes, 0 = no. $RR = \sum a_i N_{0i} / \sum (1 - a_i) N_{1i}$.

The Case-Time-Control Design

One of the limitations of the case-crossover design is the assumption of an exposure time trend. 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.⁴⁶ 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.

The approach is illustrated with data from the Saskatchewan Asthma Epidemiologic Project,¹⁵ a study conducted to investigate the risks associated with the use of inhaled β -agonists in the treatment of asthma. Using a cohort of 12 301 asthmatics followed during 1980–87, 129 cases of fatal or near-fatal asthma and 655 controls were identified. The amount of β -agonist used in the year prior to the index date was used for exposure. Table 44.3 displays the data comparing low (12 or fewer canisters per year) with high (more than 12) use of β -agonists. The crude odds ratio for high β -agonist use is 4.4 (95% CI: 2.9–6.7). Adjustment for all available markers of severity, such as oral corticosteroids and prior asthma hospitalizations as confounding factors, lowers the odds ratio to 3.1 (95% CI: 1.8–5.4), the “best” estimate one can derive from these case-control data using conventional tools.

To apply the case-time-control design, exposure to β -agonists was obtained for the 1-year current

period and the 1-year reference period prior to the current period. First, a case-crossover analysis is performed using the discordant subjects among the 129 cases, namely the 29 who were current high users of β -agonists and low users in the reference period and the nine cases who were current low users of β -agonist and high users previously. This analysis is repeated for the 655 controls, of which there were 90 discordant in exposure, that is 65 were current high users of β -agonists and low users in the reference period and 25 were current low users of β -agonists and high users previously. The case-time-control odds ratio, using these discordant pairs frequencies for a pair-matched analysis, is given by $(29/9)/(65/25) = 1.2$ (95% CI: 0.5–3.0). This estimate, which excludes the effect of unmeasured confounding by disease severity, indicates a minimal risk for these drugs.

The case-time-control approach 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 strict assumptions so caution is recommended in its use.^{40,47}

Drug Database Designs

One of the distinguishing features of pharmacoepidemiology is the use of computerized administrative health databases to answer research questions reliably with sufficient rapidity. The usual urgency of concerns related to drug safety

Table 44.3. 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		OR	Adjusted	
	High	Low	High	Low		OR	95% CI
Current β -agonist use	93	36	241	414	4.4	3.1	1.8–5.4
Discordant* use (case crossover)	29	9			3.2		1.5–6.8
Discordant* use (control crossover)			65	25	2.6		1.6–4.1
Case time control	29	9	65	25	1.2		0.5–3.0

* Discordant from exposure level during reference time period.

makes these databases essential to perform such risk assessment studies. Some databases contain only information on prescriptions dispensed to patients, and no outcome information on disease diagnoses, hospitalizations, or vital status. These standalone prescription drug databases are more numerous and usually more easily accessible than the fully linked databases. They have been the object of recent methodological developments.

A technique that was developed specifically for the drug databases is prescription sequence analysis.⁴⁸ 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 with a drug history 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, this number of subjects should be proportional to the duration of use of drug A relative to the total period of observation. This extremely rapid method of assessing the association between drug A and drug B is assessed for its random error with a Monte Carlo simulation analysis. This technique was applied to assess whether the anti-vertigo or anti-migraine drug flunarizine (drug A) causes mental depression, as measured by the use of anti-depressant drugs (drug B). The authors found that the number of patients starting on anti-depressant drugs during flunarizine use was in fact lower than expected.⁴⁸ They thus concluded, using this rapid approach based solely on drug prescription data, that this drug probably does not cause mental depression. An extension of prescription sequence analysis, called prescription sequence symmetry analysis, was recently proposed.⁴⁹ Using a population of new users of either drug A or B, this approach compares the number of subjects who used drug A before drug B to that who used B before A. Under the null hypothesis, this distribution should be symmetrical and the numbers should be equal. It has recently been applied to the question of screening for drug-related dyspepsia.⁵⁰

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.⁵¹ For example, a case-control study conducted in New Zealand found that fenoterol, a β -agonist bronchodilator used to treat asthma attacks, was associated with an increased risk of death from asthma.⁵² Using a prescription drug database, Petri *et al.* found that severe asthmatics, as deemed from their use of other asthma medications prescribed for severe forms of the disease, were in fact channeled to fenoterol, probably because fenoterol was felt by prescribers to be a more potent bronchodilator than other β -agonists.⁵³ This phenomenon of channeling can be assessed rapidly in such databases, provided medications can be used as proxies for disease severity. This approach can be subject to bias, however, as it has been used with cross-sectional designs that cannot differentiate the directionality of the association.

An application of channeling using a longitudinal design was recently presented.⁵⁴ It indicated that channeling can vary according to the timing of exposure, namely that disease severity was not associated with first-time use of a drug, but subsequently severe patients were more likely to be switched to that drug. This type of research into patterns of drug prescribing and drug use can be very useful in understanding the results of case-control studies with limited data on drug exposures and subject to confounding by indication.

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 three recently developed methodologic approaches that facilitate the conduct of research in pharmacoepidemiology. First, we presented three strategies of sampling within a large cohort, as alternatives to analyzing the full cohort, namely the nested case-control and

case-cohort techniques, as well as the extreme-stratified nested case-control approach, which has been proposed specifically to circumvent a restriction inherent in the nested case-control strategy. Future developments in this area will provide user-friendly tools to facilitate the estimation of risk difference or excess risk measures, in addition to the standard risk ratio measures routinely produced by these sampling schemes. As well, we can anticipate statistical models of analysis which take into account two or more time axes simultaneously, for example calendar time and disease duration, so that we do not have to resort to schemes based on cohort stratification. Finally, techniques to provide the optimal number of controls for each case are being devised.⁵⁵

We also described methods useful in dealing with confounders when data on this factor are only available in the cases of a case-control study or in a sample of subjects. Development in this area is crucial in view of the limitations of databases. It should be directed to the situation of multiple confounders and should address the issue of effect modification, which eludes current techniques.

Lastly, we described the case-crossover and case-time-control designs, alternatives to the classical case-control design, which are valuable when dealing with the study of transient drug effects on the risk of acute adverse events. Extensions and refinements of these designs should address their assumptions, as well as modifications for chronic effects and latent events. We also introduced novel techniques devised for standalone prescription drug databases. Such innovative methods should be given priority in view of the importance of these databases in pharmacoepidemiology.

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.

ACKNOWLEDGEMENT

Samy Suissa is the recipient of a Senior Scientist award from the Medical Research Council (MRC) of Canada.

REFERENCES

1. Thun MJ, Namboodiri MM, Heath CW II. Aspirin use and reduced risk of fatal colon cancer. *New Engl J Med* 1991; **23**: 1593-6.
2. Stampfer MJ, Willett WC, Colditz GA, Speizer FE, Hennekens CH. A prospective study of past use of oral contraceptive agents and risk of cardiovascular diseases. *New Engl J Med* 1988; **319**: 1313-7.
3. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *New Engl J Med* 1971; **284**: 878-81.
4. McFarlane MJ, Feinstein AR, Horwitz RI. Diethylstilbestrol and clear cell vaginal carcinoma: reappraisal of the epidemiologic evidence. *Am J Med* 1986; **81**: 855-63.
5. Aelony Y, Laks M, Beall G. An electrocardiographic pattern of acute myocardial infarction associated with excessive use of aerosolized isoproterenol. *Chest* 1975; **68**: 107-12.
6. Boivin JF, Hutchison GB, Lyden M, Godbold J, Chorosh J, Schottenfeld D. Second primary cancers following treatment of Hodgkin's disease. *J Natl Cancer Inst* 1984; **72**: 233-41.
7. Boivin JF, Hutchinson GB, Lubin JH, Mauch P. Coronary artery disease mortality in patients treated for Hodgkin's disease. *Cancer* 1992; **69**: 1241-7.
8. Vessey MP, Villard-Mackintosh L, McPherson K, Yeates D. Mortality among oral contraceptive users: 20 year follow up of women in a cohort study. *Br Med J* 1989; **299**: 1487-9.
9. Rothman KJ. *Modern Epidemiology*. Boston: Little, Brown, 1986.
10. Miettinen OS. *Theoretical epidemiology: principles of occurrence research in medicine*. New York: Wiley, 1985.
11. Suissa S, Spitzer WO, Abenhaim L, Downey W, Gardiner RJ, Fitzgerald D. Risk of death from human insulin. *Pharmacoepidemiol Drug Safety* 1992; **1**: 169-75.
12. Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973; **29**: 479-86.
13. Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis appraisal by application to asbestos mining. *J R Stat Soc A* 1977; **140**: 469-91.
14. Lubin JH, Gail MH. Biased selection of controls for case-control analyses of cohort studies. *Biometrics* 1984; **40**: 63-75.

15. Spitzer WO, Suissa S, Ernst P, *et al.* The use of β -agonists and the risk of death and near death from asthma. *New Engl J Med* 1992; **326**: 501–6.
16. Breslow NE, Day NE. *Statistical Methods in Cancer Research. Vol. I. The Analysis of Case-Control Studies (IARC Sci. Publ. No. 32)*. Lyon, 1980.
17. EGRET. Seattle, WA: Statistical and Epidemiology Research Corporation, 1990.
18. Breslow NE, Day NE. *Statistical Methods in Cancer Research. Vol. II. The Design and Analysis of Cohort Studies (IARC Sci. Publ. No. 82)*. Lyon, 1987.
19. Suissa S, Hemmelgarn B, Spitzer WO, *et al.* The Saskatchewan oral contraceptive cohort study of oral contraceptive use and cardiovascular risks: research protocol. *Pharmacoepidemiol Drug Safety* 1993; **2**: 33–49.
20. Suissa S, Edwardes MD, Boivin JF. External comparisons in nested case-control designs. *Epidemiology* 1997; **9**: 72–8.
21. Hutchison GB. Leukemia in patients with cancer of the cervix uteri treated with radiation. A report covering the first 5 years of an international study. *J Natl Cancer Inst* 1968; **40**: 951–82.
22. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; **73**: 1–11.
23. Bergkvist L, Adami HO, Persson I, Hoover R, Schairer C. The risk of breast cancer after estrogen and estrogen-progestin replacement. *New Engl J Med* 1989; **321**: 293–7.
24. 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.
25. Moulton LH, Wolff MC, Brennehan G, Santosham M. Case-cohort analysis of case-coverage studies of vaccine effectiveness. *Am J Epidemiol* 1995; **142**: 1000–6.
26. Strom BL, Schinnar R, Bilker WB, Feldman H, Farrar JT, Carson JL. Gastrointestinal tract bleeding associated with naproxen sodium vs ibuprofen. *Arch Intern Med* 1997; **157**: 2626–31.
27. EPICURE. Risk regression and data analysis software. Seattle, WA: Hirosoft International Corporation, 1991.
28. Wacholder S. Practical considerations in choosing between the case-control and nested case-control designs. *Epidemiology* 1991; **2**: 155–8.
29. Wacholder S, Boivin JF. External comparisons with the case-cohort design. *Am J Epidemiol* 1987; **126**: 1198–209.
30. Hemmelgarn B, Suissa S, Boivin JF, Huang A, Pinard G. Benzodiazepine use and the risk of motor vehicle crashes in the elderly [see editorial] *J Am Med Assoc* 1997; **278**: 27–31.
31. Ray WA, Griffin MR. Use of Medicaid data for pharmacoepidemiology. *Am J Epidemiol* 1989; **129**: 837–49.
32. Griffin MR, Ray WA, Schaffner W. Nonsteroidal anti-inflammatory drug use and death from peptic ulcer disease in the elderly. *Ann Intern Med* 1988; **109**: 359–63.
33. Jick H, Watkins RN, Hunter JR, Dinan BJ, Madsen S, *et al.* Replacement estrogens and endometrial cancer. *New Engl J Med* 1979; **300**: 218–22.
34. Suissa S. Statistical methods in pharmacoepidemiology: principles in managing error. *Drug Safety* 1991; **6**: 381–9.
35. Suissa S, Edwardes M. Adjusted odds ratios for case-control studies with missing confounder data in controls. *Epidemiology* 1997; **8**: 275–80.
36. Suissa S, Hemmelgarn B, Blais L, Ernst P. Bronchodilators and acute cardiac death. *Am J Resp Crit Care Med* 1996; **154**: 1598–602.
37. 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.
38. Bilker WB, Berlin JA, Gail MH, Strom BL. An efficient design for verifying disease outcome status in large cohorts with rare exposures and low disease rates. *Stat Med* 1999; **18**: 3021–36.
39. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991; **133**: 144–53.
40. Greenland S. Confounding and exposure trends in case-crossover and case-time-control design. *Epidemiology* 1996; **7**: 231–9.
41. Ray WA, Fought RL, Decker MD. Psychoactive drugs and the risk of injurious motor vehicle crashes in elderly drivers. *Am J Epidemiol* 1992; **136**: 873–83.
42. Sturkenboom MC, Middelbeek A, de Jong van den Berg LT, van den Berg PB, Stricker BH, Wesseling H. Vulvo-vaginal candidiasis associated with acitretin. *J Clin Epidemiol* 1995; **48**: 991–7.
43. Meier CR, Jick SS, Derby LE, Vasilakis C, Jick H. Acute respiratory-tract infections and risk of first-time acute myocardial infarction. *Lancet* 1998; **351**: 1467–71.
44. Barbone F, McMahon AD, Davey PG, Morris AD, Reid IC, McDevitt DG, MacDonald TM. Association of road-traffic accidents with benzodiazepine use. *Lancet* 1998; **352**: 1331–6.
45. Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation [Review]. *Am J Epidemiol* 1996; **143**: 1165–73.
46. Suissa S. The case-time-control design. *Epidemiology* 1995; **6**: 248–53.

47. Suissa S. The case-time-control design: further assumptions and conditions. *Epidemiology* 1998; **9**: 441–5.
48. 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.
49. Hallas J. Evidence of depression provoked by cardiovascular medication: a prescription sequence symmetry analysis. *Epidemiology* 1996; **7**: 478–84.
50. Hallas J, Butzer P. Screening for drug related dyspepsia: an analysis of prescription symmetry. *Eur J Gastroenterol Hepatol* 1998; **10**: 27–32.
51. Urquhart J. ADR crisis management. *Scrip* 1989; **1388**: 19–21.
52. Crane J, Pearce N, Flatt A, Jackson R, Ball M, Burgess C, *et al.* Prescribed fenoterol and death from asthma in New Zealand 1981–1983: case-control study. *Lancet* 1989; **1**: 917–22.
53. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med* 1991; **10**: 577–81.
54. 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.
55. Miettinen OS. Principles of epidemiologic research I: study design. *Course Notes*. McGill University, Montreal, Canada, 1993.

Part V
CONCLUSION

45

The Future of Pharmacoepidemiology

BRIAN L. STROM

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

We should all be concerned about the future because we will have to spend the rest of our lives there.

Charles Franklin Kettering, 1949

Speculating about the future is risky and, some might say, foolish. Nevertheless, based on events occurring today, in many ways the future of pharmacoepidemiology is probably apparent. Interest in the field is exploding, as is realization of what it can contribute. As science, industry, and government become increasingly international, so does the field of pharmacoepidemiology. The number of individuals attending the annual *International Conference on Pharmacoepidemiology* has increased from 50 to over 500. The International Society for Pharmacoepidemiology, only a decade old, has grown to over 1200 members from 57 countries. Many national societies are being formed as well. The journal *Clinical Pharmacology and Therapeutics*, the major US academic clinical pharmacology journal, is actively soliciting pharmacoepidemiology manuscripts, as is the *Journal of Clinical Epidemiology*. There are also at least two journals (down from five) which explicitly consider

themselves pharmacoepidemiology journals: the *Journal of Pharmacoepidemiology* and *Pharmacoepidemiology and Drug Safety*, the latter being the official journal of the International Society for Pharmacoepidemiology and the European Society for Pharmacovigilance. The number of individuals seeking to enter the field is rapidly increasing. The number of courses in pharmacoepidemiology is increasing, in schools of medicine, public health, and pharmacy. In fact, a decade ago, a paper expressed the belief that all doctor of pharmacy students should receive some instruction in pharmacoepidemiology.¹ While in past years the single summer short course in pharmacoepidemiology had to be cancelled due to lack of US enrollees, more recently the University of Michigan summer course in pharmacoepidemiology has attracted 10% of the entire program, the Boston summer program established a course at the identical time, McGill University provides well attended short courses in pharmacoepidemiology all around the world, and many other *ad hoc* short courses are given as well. Regulatory bodies have expanded their internal programs. The number of pharmaceutical companies forming their own

pharmacoepidemiology units has also increased, along with their funding of academic units. Requirements that a drug be proven to be cost-effective (see Chapter 35) have been added to many national healthcare systems, provincial healthcare systems, or managed care organizations, either to justify reimbursement or even to justify drug availability. Drug utilization review is being widely applied, and hospitals are becoming mini pharmacoepidemiology laboratories (see Chapter 31).

Thus, although some important obstacles remain, the future of the field looks extraordinarily bright. In this chapter, I will briefly risk giving my views on where I see the future of the field heading. I will follow the format of Part II of the book, that is I will explore the future of the field from the perspective of academia, then from the perspective of the pharmaceutical industry, then from the perspective of a regulatory agency, and last from the perspective of the law.

THE VIEW FROM ACADEMIA

SCIENTIFIC DEVELOPMENTS

Methodologic Advances

Methodologically, it appears likely that the number of approaches available for performing pharmacoepidemiology studies will continue to increase. Each of the methodologic issues discussed in Part IV can be expected to be the subject of more development. The future is likely to see ever more sophisticated ways of performing and analyzing classical epidemiologic studies, as the field of epidemiology continues to expand and develop. Some of these new techniques will, of course, be useful to investigators in pharmacoepidemiology (see Chapters 43 and 44). Some new methodologies likely to emerge as of importance to pharmacoepidemiology over the next few years will be the use of instrumental variables, neural networks, propensity scores, recursive partitioning, and sensitivity analysis. In addition, I think we will see increasing use of pharmacoepidemiology

studies and approaches, to improve the conduct of clinical trials.

The analysis of spontaneous reports of adverse drug effects, and the process of deciding whether a drug really caused any given adverse outcome observed in a case report, can be expected to continue to develop (see Chapters 10, 11, and 32). The need for new methods to screen for adverse drug effects is clear. It is likely that systems for screening for unexpected drug effects will emerge, both using spontaneous reporting systems and otherwise.

We can expect to see increasing cross-fertilization between new developments in clinical pharmacology and pharmacoepidemiology, with more use of therapeutic drug monitoring in pharmacoepidemiology studies (see Chapter 4) and more input from pharmacoepidemiologists into policy questions about drug availability (see Chapter 27). More emphasis will be placed on studies of predictors of who is at risk of adverse reactions, rather than just whether a drug causes an adverse reaction,² as, if risk factors for an adverse reaction can be better understood, early in the course of a crisis, a drug is more likely to be able to be repositioned in the market, rather than withdrawn, saving good drugs.

With the development of the techniques of molecular biology, and their application to the study of pharmacogenetics, recent years have seen exciting developments in the ability of researchers to determine the genetic basis for adverse drug reactions previously thought to be idiosyncratic.³ This has evolved from earlier measures of slow drug metabolism as a cause of adverse reactions⁴ to more recent molecular genetic markers.⁵ This has been aided by the development of new methods to collect DNA, making population based genetic studies feasible. There are even studies under way now that could be called molecular pharmacoepidemiology, e.g., the author is leading a pair of large scale population-based case-control studies examining the genetic basis of susceptibility to hormone-induced cancers, using cheek swabs to collect DNA specimens.

Advances can also be expected in the identification of drug exposures in human tissue. Blood and urine have long been available for this purpose,

but detection is largely limited to the interval shortly following exposure. For case-control studies in particular, where the outcome is typically identified some time after exposure, such sampling is of no use for detecting earlier exposures. In recent years, however, researchers have explored the usefulness of other tissues in which drugs or their metabolites may persist and accumulate, such as hair.

Future years are likely to see much more of this cross-fertilization between pharmacoepidemiology and the new biology. By analogy with the process of screening for rubella susceptibility or for genetic diseases, it is not unreasonable to look forward to a time when patients can be screened for drug-specific genetic polymorphisms which place them at particular risk for having an adverse reaction. Information of this kind has obvious usefulness in selecting (and avoiding) specific drugs for the treatment of specific individuals. We could easily see individual pharmacogenetic studies added to the process of following up on spontaneous reports of adverse reactions.

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 sophisticated. As the healthcare industry becomes more sensitive to the possibility of overutilization, underutilization, and inappropriate utilization of drugs, and the risks associated with each, one would expect to see an increased frequency of and sophistication in drug utilization review programs. This is likely to be the case especially with studies of antibiotic misuse, as society becomes ever more concerned about the evolution of organisms resistant to our currently available drugs. Drug utilization studies will continue to increase (see Chapter 29), and drug utilization review programs will expand enormously, as attempts to improve physician prescribing continue to expand (see Chapter 31). In fact, drug utilization review programs are already required in every US hospital, are required of every US state Medicaid program, have voluntarily been initiated by many US Health

Maintenance Organizations, and are likely to be written into the law establishing the new US Medicare drug benefit, should it be passed.

The US Joint Commission on Accreditation of Healthcare Organizations is revolutionizing US hospital pharmacoepidemiology through its new standards requiring adverse reaction monitoring and drug use evaluation programs in every hospital^{6,7} (see Chapters 12, 29, 30, and 31). The initial Joint Commission plan for an Indicator Measurement System⁸ has now evolved into its ORYX system,⁹ as a means of collecting ongoing quantitative data on a hospital's performance. Hospitals are 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.¹⁰ The US Health Care Financing Administration, the branch of the federal government that manages the Medicaid and Medicare programs, is now proposing new rules, which will require increasing attention in hospitals to drug safety issues.¹¹ This is likely to revolutionize hospital pharmacoepidemiology in the US, and it is likely that the next edition of this book will require a separate, new chapter describing the results of these initiatives.

Interest in the new field of "pharmacoeconomics," that is the application of the principles of health economics to the study of drug effects, is continuing to explode (see Chapter 35). Society is realizing that the actual cost of drugs is a very minor part of their economic impact, and that their beneficial and harmful effects are vastly more important. On the other hand, more countries and insurance programs within countries are requiring economic justification before permitting reimbursement for a drug. The number of studies exploring this is increasing dramatically. As the methods of pharmacoeconomics become increasingly sophisticated, and its applications clear, this could be expected to continue to be a fruitful and popular field of inquiry.

More studies of beneficial drug effects, particularly of drug effectiveness, can be expected, as the field becomes more aware that they are possible (see Chapter 34). This is being aided by the rapid increase in the use of propensity scores to adjust

for confounding by indication, although often placing more confidence in that technique than is warranted, not recognizing that its use is still dependent on one's ability to *measure* the true determinants of the indication (see Chapter 34).

We will also see more use of pharmacoepidemiology approaches prior to drug marketing, e.g., in order to understand the adverse events that one can expect to see in patients to be treated with a new drug (see Chapters 7 and 28).

Recent years have seen an explosion in the use of herbal medications, worldwide. These are essentially pharmaceuticals that are being marketed 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. As such, it is quite likely that the next few years will see a comparable set of safety disasters associated with their use, and pharmacoepidemiologists will be turned to, to help evaluate them.

Research interest in the whole topic of patient noncompliance with prescribed drug regimens goes back to about 1960, but little could be done until about 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. One was to use very low doses of a very long half-life agent, phenobarbital, as a chemical marker; a single measurement of phenobarbital in plasma is indicative of aggregate drug intake during the prior two weeks.¹² The other has been to incorporate time-stamping microcircuitry into pharmaceutical packages, which creates a timed, dated record of each time maneuvers are made with the package that are necessary to remove a dose of drug.¹³ 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.

Finally, 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 methodologies, and increasing use of techniques from laboratory sciences, as described above. The former will allow us to use predictive modeling to study, from a population perspective, who is statistically most likely to benefit from a drug, and who is at greatest risk of an adverse outcome. The latter will enable us, as noted above, to test individuals genetically, to determine their likely responses to drug therapy. From a drug testing point of view, these developments will allow researchers to target for enrollment into their studies those subjects most likely to succeed with a drug. From a clinical perspective, it will enable healthcare prescribers to individualize our choice of therapies.

Logistical Advances

Logistically, with the increased computerization of society in particular and the health care industry, in particular, and the increased emphasis on using computerized databases for pharmacoepidemiology¹⁴ (see Part III), some data resources will disappear (e.g., The Rhode Island Drug Use Reporting System and the inpatient databases discussed in prior editions of this book have disappeared), and a number of new computerized databases will undoubtedly emerge as major resources for pharmacoepidemiology research (e.g., see Chapters 16–24). Others may well follow. A few have even disappeared and then re-emerged (see Chapters 20 and 23). The importance of these databases to pharmacoepidemiology is now clear: they are able to address, quickly and relatively inexpensively, questions about drug effects that require large sample sizes, with excellent quality data on drug exposures.

Nevertheless, as the field continues to evolve toward the increased use of databases, it is important to keep in mind the vital importance of resources that collect their data *de novo*. Each approach to pharmacoepidemiology has its advantages and its disadvantages, as described in Chapters 10–25. No approach is ideal, and often a number of complementary approaches must be used to answer any given research question. To address some of the problems of the databases, we

must always maintain the ability to test their findings using data collected in *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 used is the network of regional and national poison control centers¹⁰ (also see Chapter 24). 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 other new types of opportunity for research will emerge. For example, as the US discusses the expansion of Medicare, its health program for the elderly, it appears likely that one component will be the reimbursement of individuals for the costs of their prescription medications. This should generate an enormous new data resource that potentially could be useful for pharmacoepidemiology research. Outside the US, as well, many different opportunities to form databases are being sought (e.g., see Chapters 20–24). There is also an increased interest in the importance of pharmacoepidemiology in the third world. Many third world countries spend a disproportionate amount of their health care resources on drugs,¹⁵ yet these drugs are being used inappropriately.¹⁶ There have been a number of initiatives in response to this, including the World Health Organization's development of its list of "essential drugs."^{17,18} The International Clinical Epidemiology Network (INCLIN), with the support of The Rockefeller Foundation, embarked on a major initiative to develop more pharmacoepidemiology interest and expertise in the third world.¹⁹

FUNDING

For a number of years, academic pharmacoepidemiology suffered from limited research funding opportunities. With the increasing interest in the field, this seems to be in the process of changing, in the US at least. Much more industry funding is available, as industry sees the need for the field (see below). It appears this is likely to increase, especially as the FDA rebuilds its own

pharmacoepidemiology program, and more often requires "phase IV" studies. This will be particularly true, if these new "phase IV" studies are used to shorten phase III and permit earlier drug marketing, as has been proposed for drugs used to treat life threatening illnesses and has been operationalized in selected situations, e.g., zidovudine.²⁰ In fact, one could argue that postmarketing pharmacoepidemiology studies should be performed for all newly marketed drugs that are used for chronic diseases, or expected to be either novel or blockbusters, because of the risks which these present.

There is, of course, a risk associated with academic groups becoming too dependent on industry funding, both in terms of independence and credibility. Fortunately, in the US FDA's funding for extramural pharmacoepidemiology contracts and cooperative agreements has continued, although it has not been increased, and therefore the number of funded programs has in fact been reduced. The US Agency for Health Care Policy and Research (AHCPR) has begun to fund pharmacoepidemiology research as well, as part of an initiative in pharmaceutical outcomes research. In particular, the AHCPR CERT program (Centers for Education and Research in Therapeutics) appears particularly promising, to begin to provide federal support for ongoing pharmacoepidemiology activities.

Even the National Institutes of Health (NIH) has begun to fund pharmacoepidemiology projects more often, including a small program in pharmacology in the National Institute on Aging. NIH is the logical source for most 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 support for our field. It remains the most appropriate source of such support: it is the institute intended to fund projects that cross organ systems, and it is the institute that

funds clinical pharmacology research. However, over the past few years it has declined to fund any epidemiology research, as it is focusing much of its resources on molecular biology. This remains a problem for the field of pharmacoepidemiology, a problem that badly needs to be addressed. In the meantime, NIH funding is available now if one tailors a project to fit an organ system or the priorities of another one of the individual institutes.

If the US government begins to pay for drugs as part of Medicare, and therefore becomes concerned about the use of, effects of, and costs of drugs, it is possible that there will be substantial new funding for pharmacoepidemiology available.

Finally, but of critical importance, there is increasing concern about confidentiality in many countries throughout the world (see Chapter 26). The regulatory framework for human research is actively changing, in the process. As discussed elsewhere, this is already beginning to interfere with pharmacoepidemiology research, 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 manpower. There is a crying need for more people in the field, with opportunities available in academia, industry, and FDA. 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.

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, which funds training programs in clinical pharmacology, has refused to support epidemiology. This results in the dependence of interdisciplinary training on nonfederal sources of funds. There are a few institutions now capable of carrying out such training, for example schools of public health that have a faculty member interested in pharmacoepidemiology and institutions that have developed clinical epidemiology training programs with support from, for example, the Andrew W. Mellon Foundation and The Rockefeller Foundation. Young scientists interested in receiving training in pharmacoepidemiology, however, can only do so if they happen to qualify for support from one of these other programs. No ongoing support is normally available from these programs for training in pharmacoepidemiology *per se*. Attempts are under way to try to address this, however, primarily through the leadership and generosity of selected pharmaceutical manufacturers. 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 of what was said above about the future of pharmacoepidemiology scientifically, as it relates to academia, obviously relates to industry, as well. The utility of pharmacoepidemiology to industry has become apparent (see Chapters 5 and 7). In addition to being useful for exploring the effects of their drugs, manufacturers are beginning to realize that the field can not only find problems, but it can document drug safety. An increasing number of manufacturers are mounting pharmacoepidemiology studies prophylactically, to have data available in advance when crises occur. Proper practice would argue for phase IV observational studies for all newly marketed

drugs used for chronic diseases, and all drugs expected to be either novel or blockbusters, because of the risks these present. Pharmacoepidemiology also can be used for quantitating beneficial drug effects (see Chapter 34) and even for marketing purposes, both marketing research and direct marketing efforts. Perhaps most importantly, pharmacoepidemiology studies can be used to protect the major investment made in developing a new drug against inappropriate accusations of adverse effects.

In light of these advantages, most major pharmaceutical firms have formed their own in-house pharmacoepidemiology units. Of course, this then means that industry confronts and, in fact, exaggerates the problem of insufficient trained personnel, which was described above. Many pharmaceutical companies are also increasing their investment in external pharmacoepidemiology resources, so they will be available when crises arise. All of this is likely to continue. The risk of the growth of pharmacoepidemiology for industry is the generation of an increased number of false signals about harmful drug effects. This can only be addressed by having adequately trained individuals in the field to minimize this risk, and by having personnel and resources available to address these questions quickly, responsibly, and effectively, if or when they are raised.

THE VIEW FROM REGULATORY AGENCIES

It appears that the role of pharmacoepidemiology in regulatory agencies is also expanding (see Chapter 8). Again, all of what was said above about the future of pharmacoepidemiology scientifically, as it relates to academia, obviously relates to regulatory agencies, as well. In addition, there have been a large series of major drug problems that have occurred, detailed throughout this book. Many of these resulted in the removal of the drug 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.²⁰ 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 in the Agency for Healthcare Policy and Research, the Centers for Disease Control, and the National Institutes of Health, and there is increasing debate about the need to develop an independent new Center for Drug Surveillance.

THE VIEW FROM THE LAW

Finally, as explored in Chapter 9, the importance of pharmacoepidemiology to the law has also been increasing. The number of lawsuits relating to pharmacoepidemiology questions is very large. There are an increasing number of drugs on the market, an increasing sensitivity to the adverse effects they can have, and an increasing awareness of the legal system's ability to obtain substantial remuneration for those who suffered from those adverse effects. Thus, I would predict that the issues in future litigation will shift, from say pelvic inflammatory disease caused by intrauterine devices to the cardiac valvular effects of diet drugs. However, it is probable that the number of issues subject to litigation will only increase. The interest in the field and the need for more true experts in the field will, therefore, increase accordingly.

CONCLUSION

There are no really "safe" biologically active drugs. There are only "safe" physicians.

Harold A. Kaminetzsky, 1963

All drugs have adverse effects. Pharmacoepidemiology will never succeed in preventing them; it can only detect them and, thereby, educate physicians so that they can be better prescribers. The result will be better for industry and academia but, most importantly, for public health. The next

“thalidomide disaster” cannot be prevented by pharmacoepidemiology. However, pharmacoepidemiology can minimize its adverse public health impact. At the same time, it can rationalize the use of drugs that belong on the market, and protect against the inappropriate 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 former accentuated and the latter solved.

REFERENCES

- Whitney HA. Assimilating pharmacoepidemiology into education and practice. *Drug Intell Clin Pharm* 1987; **21**: 993.
- 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.
- Spielberg SP. Idiosyncratic drug reactions: interaction of development and genetics. *Semin Perinatol* 1992; **16**: 58–62.
- Woodsley 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. *New Engl J Med* 1978; **298**: 1157–9.
- 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.
- 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.
- Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. *J Am Med Assoc* 1991; **266**: 2847–51.
- Nadzam DM, Turpin R, Hanold LS, White RE. Data-driven performance improvement in health care: The Joint Commission’s Indicator Measurement System (IMSystem). *Joint Commission J Qual Improv* 1993; **19**: 492–500.
- Joint Commission on Accreditation of Healthcare Organizations. ORYX: the next evolution in accreditation. *Nurs Management* 1997; **28**: 49–54.
- Bates DW, Leape LL, Cullen DJ, Laird N, Petersen LA, Teich JM, Burdick E, Hickey M, Kleeffeld S, Shea B, Vander Vliet M, Seger DL. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *J Am Med Assoc* 1998; **280**: 1311–6.
- Department of Health and Human Services. Health Care Financing Administration. 42 CFR Parts 416, 482, 485, and 489. *Fed Register* 1997; **62**, 66 726–57.
- Pullar T, Feely M. Problems of compliance with drug treatment: new solutions? *Pharm J* 1990; **245**: 213–5.
- Urquhart J. The electronic medication event monitor—lessons for pharmacotherapy. *Clin Pharmacokinetics* 1997; **32**: 345–56.
- Strom BL, Carson JL. Use of automated databases for pharmacoepidemiology research. *Epidemiol Rev* 1990; **12**: 87–107.
- 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): 11–7.
- Lunde PKM. WHO’s programme on essential drugs. Background, implementation, present state and perspectives. *Dan Med Bull* 1984; **31** (suppl 1): 23–7.
- Howard NJ, Laing RO. Changes in the World Health Organization essential drug list. *Lancet* 1991; **338**: 743–5.
- Guess HA, Goldsmith C, Henry D, Strom BL. Pharmacoepidemiology as a focus for clinical epidemiology in developing countries. *J Clin Epidemiol* 1991; **44** (suppl 2): 101S–5S.
- Young FE. The role of the FDA in the effort against AIDS. *Public Health Rep* 1988; **103**: 242–5.

Appendix A

Sample Size Tables

Table A1. Sample sizes for cohort studies^a

Incidence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	2.5	2.5
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		
0.00005	394 133	557 684	1 261 219	5 885 657	7 567 179	2 101 980	630 585	326 965		
0.0001	197 060	278 832	630 585	2 942 699	3 783 376	1 050 923	315 268	163 467		
0.0005	39 401	55 751	126 078	588 332	756 333	210 078	63 015	32 669		
0.001	19 694	27 865	63 015	294 037	377 953	104 973	31 483	16 320		
0.005	3 928	5 557	12 564	58 600	75 249	20 888	6 257	3 240		
0.01	1 957	2 769	6 257	29 170	37 411	10 378	3 104	1 605		
0.05	381	538	1 212	5 627	7 140	1 969	582	297		
0.10	184	259	582	2 684	3 357	918	266	133		
0.15	118	166	372	1 703	2 095	568	161	79		
0.20	85	120	266	1 212	1 465	393	109	52		
0.25	65	92	203	918	1 086	287	77	35		
0.30	52	73	161	722	834	217	56	24		
0.35	43	60	131	582	654	167	41	16		
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 A1. Sample sizes for cohort studies^a (continued)

Incidence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	1 051 034	756 742	583 904	394 133	211 445	142 727	61 134	22 318		
0.00005	210 189	151 334	116 768	78 816	42 280	28 538	12 220	4 458		
0.0001	105 083	75 657	58 376	39 401	21 135	14 264	6 106	2 225		
0.0005	20 999	15 117	11 662	7 870	4 219	2 845	1 215	439		
0.001	10 488	7 549	5 823	3 928	2 104	1 418	603	216		
0.005	2 080	1 495	1 152	775	412	276	114	37		
0.01	1 028	738	568	381	201	133	53	15		
0.05	188	133	101	65	32	19	4	—		
0.10	82	57	42	26	10	4	—	—		
0.15	47	32	23	13	—	—	—	—		
0.20	30	19	13	6	—	—	—	—		
0.25	19	12	7	—	—	—	—	—		
0.30	12	6	—	—	—	—	—	—		
0.35	7	—	—	—	—	—	—	—		
0.40	—	—	—	—	—	—	—	—		
0.45	—	—	—	—	—	—	—	—		
0.50	—	—	—	—	—	—	—	—		
0.55	—	—	—	—	—	—	—	—		
0.60	—	—	—	—	—	—	—	—		
0.65	—	—	—	—	—	—	—	—		
0.70	—	—	—	—	—	—	—	—		
0.75	—	—	—	—	—	—	—	—		
0.80	—	—	—	—	—	—	—	—		
0.85	—	—	—	—	—	—	—	—		
0.90	—	—	—	—	—	—	—	—		
0.95	—	—	—	—	—	—	—	—		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.10$ (power = 90%), control : exposed ratio = 1 : 1. The sample size listed is the number of subjects needed in the exposed group. An equivalent number would be included in the control group.

Table A2. Sample 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	2.5	2.5
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		
0.00005	305 803	430 714	965 093	4 455 806	5 629 566	1 552 830	460 550	236 696		
0.0001	152 896	215 349	482 527	2 227 804	2 814 625	776 367	230 258	118 337		
0.0005	30 570	43 057	96 475	445 402	562 673	155 196	46 024	23 651		
0.001	15 280	21 521	48 218	222 602	281 179	77 550	22 994	11 815		
0.005	3 047	4 292	9 613	44 362	55 984	15 433	4 571	2 346		
0.01	1 518	2 138	4 787	22 082	27 834	7 668	2 268	1 163		
0.05	295	415	927	4 258	5 315	1 456	426	216		
0.10	142	200	444	2 030	2 500	680	196	97		
0.15	91	128	283	1 287	1 561	421	119	58		
0.20	66	92	203	916	1 092	291	80	38		
0.25	50	70	155	693	811	214	57	26		
0.30	40	56	123	545	623	162	42	18		
0.35	33	46	100	439	489	125	31	12		
0.40	27	38	82	359	388	97	22	8		
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 A2. Sample sizes for cohort studies^a (continued)

Incidence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	755 529	540 883	415 381	278 329	147 626	99 000	41 938	15 197		
0.00005	151 093	108 167	83 068	55 659	29 520	19 795	8 384	3 036		
0.0001	75 539	54 077	41 528	27 825	14 756	9 895	4 189	1 516		
0.0005	15 095	10 805	8 297	5 558	2 946	1 974	834	300		
0.001	7 540	5 396	4 143	2 774	1 469	984	414	148		
0.005	1 496	1 069	820	548	288	192	79	26		
0.01	740	528	404	269	141	93	37	11		
0.05	136	95	72	47	23	14	3	—		
0.10	60	41	31	19	8	3	—	—		
0.15	35	23	17	9	—	—	—	—		
0.20	22	14	10	4	—	—	—	—		
0.25	14	9	5	—	—	—	—	—		
0.30	9	4	—	—	—	—	—	—		
0.35	5	—	—	—	—	—	—	—		
0.40	—	—	—	—	—	—	—	—		
0.45	—	—	—	—	—	—	—	—		
0.50	—	—	—	—	—	—	—	—		
0.55	—	—	—	—	—	—	—	—		
0.60	—	—	—	—	—	—	—	—		
0.65	—	—	—	—	—	—	—	—		
0.70	—	—	—	—	—	—	—	—		
0.75	—	—	—	—	—	—	—	—		
0.80	—	—	—	—	—	—	—	—		
0.85	—	—	—	—	—	—	—	—		
0.90	—	—	—	—	—	—	—	—		
0.95	—	—	—	—	—	—	—	—		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.10$ (power = 90%), control : exposed ratio = 2 : 1. The sample size listed is the number of subjects needed in the exposed group. Double this number would be included in the control group.

Table A3. Sample sizes for cohort studies^a

Incidence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	2.5	2.5
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		
0.00005	273 886	386 158	864 495	3 977 589	4 982 452	1 368 657	402 927	205 788		
0.0001	136 938	193 072	432 230	1 988 706	2 491 087	684 286	201 449	102 885		
0.0005	27 380	38 603	86 418	397 599	497 995	136 790	40 266	20 563		
0.001	13 685	19 294	43 192	198 711	248 859	68 352	20 118	10 272		
0.005	2 729	3 847	8 611	39 600	49 549	13 603	4 000	2 040		
0.01	1 359	1 916	4 288	19 711	24 636	6 759	1 985	1 011		
0.05	264	372	830	3 800	4 705	1 284	373	188		
0.10	127	179	398	1 811	2 213	600	171	85		
0.15	81	114	254	1 148	1 383	372	104	50		
0.20	58	82	181	817	968	257	71	33		
0.25	45	63	138	618	719	189	50	23		
0.30	36	50	109	485	552	143	37	16		
0.35	29	41	89	391	434	111	27	11		
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 A3. Sample sizes for cohort studies^a (continued)

Incidence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	653 418	465 696	356 275	237 254	124 571	83 030	34 793	12 510		
0.00005	130 673	93 131	71 248	47 445	24 910	16 602	6 955	2 499		
0.0001	65 330	46 560	35 619	23 719	12 452	8 299	3 476	1 248		
0.0005	13 055	9 303	7 117	4 738	2 486	1 656	692	247		
0.001	6 521	4 646	3 554	2 365	1 240	825	344	122		
0.005	1 294	921	703	467	244	161	66	21		
0.01	640	455	347	230	119	78	31	9		
0.05	117	82	62	40	19	12	2	—		
0.10	52	36	26	16	7	3	—	—		
0.15	30	20	14	8	—	—	—	—		
0.20	19	12	8	4	—	—	—	—		
0.25	13	7	4	—	—	—	—	—		
0.30	8	4	—	—	—	—	—	—		
0.35	4	—	—	—	—	—	—	—		
0.40	—	—	—	—	—	—	—	—		
0.45	—	—	—	—	—	—	—	—		
0.50	—	—	—	—	—	—	—	—		
0.55	—	—	—	—	—	—	—	—		
0.60	—	—	—	—	—	—	—	—		
0.65	—	—	—	—	—	—	—	—		
0.70	—	—	—	—	—	—	—	—		
0.75	—	—	—	—	—	—	—	—		
0.80	—	—	—	—	—	—	—	—		
0.85	—	—	—	—	—	—	—	—		
0.90	—	—	—	—	—	—	—	—		
0.95	—	—	—	—	—	—	—	—		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.10$ (power = 90%), control : exposed ratio = 3 : 1. The sample size listed is the number of subjects needed in the exposed group. Triple this number would be included in the control group.

Table A4. Sample sizes for cohort studies^a

Incidence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	4.0
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	—	—
0.00005	257 106	363 164	813 616	3 737 999	4 658 521	1 276 231	373 825	190 079	—	—
0.0001	128 548	181 575	406 791	1 868 916	2 329 131	638 076	186 899	95 031	—	—
0.0005	25 702	36 304	81 332	373 649	465 619	127 552	37 358	18 993	—	—
0.001	12 846	18 145	40 650	186 741	232 680	63 737	18 665	9 488	—	—
0.005	2 562	3 618	8 104	37 214	46 329	12 684	3 711	1 884	—	—
0.01	1 276	1 802	4 035	18 523	23 035	6 303	1 842	934	—	—
0.05	248	349	781	3 571	4 399	1 198	346	174	—	—
0.10	119	168	374	1 702	2 070	560	159	78	—	—
0.15	76	107	238	1 079	1 294	347	97	47	—	—
0.20	55	77	171	767	905	240	66	31	—	—
0.25	42	59	130	580	672	177	47	21	—	—
0.30	33	47	103	456	517	134	34	15	—	—
0.35	27	38	83	366	406	103	25	10	—	—
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 A4. Sample sizes for cohort studies^a (continued)

Incidence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	601 217	427 061	325 766	215 895	112 429	74 554	30 945	11 048		
0.00005	120 234	85 404	65 147	43 174	22 482	14 907	6 186	2 207		
0.0001	60 111	42 697	32 569	21 583	11 238	7 451	3 091	1 102		
0.0005	12 013	8 532	6 507	4 311	2 244	1 487	615	218		
0.001	6 000	4 261	3 249	2 152	1 119	741	306	107		
0.005	1 190	844	643	425	220	145	58	19		
0.01	589	417	318	209	107	70	27	8		
0.05	108	76	57	36	17	10	2	—		
0.10	48	33	24	15	6	2	—	—		
0.15	28	19	13	7	—	—	—	—		
0.20	18	11	8	3	—	—	—	—		
0.25	12	7	4	—	—	—	—	—		
0.30	7	—	—	—	—	—	—	—		
0.35	3	—	—	—	—	—	—	—		
0.40	—	—	—	—	—	—	—	—		
0.45	—	—	—	—	—	—	—	—		
0.50	—	—	—	—	—	—	—	—		
0.55	—	—	—	—	—	—	—	—		
0.60	—	—	—	—	—	—	—	—		
0.65	—	—	—	—	—	—	—	—		
0.70	—	—	—	—	—	—	—	—		
0.75	—	—	—	—	—	—	—	—		
0.80	—	—	—	—	—	—	—	—		
0.85	—	—	—	—	—	—	—	—		
0.90	—	—	—	—	—	—	—	—		
0.95	—	—	—	—	—	—	—	—		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.10$ (power = 90%), control : exposed ratio = 4 : 1. The sample size listed is the number of subjects needed in the exposed group. Quadruple this number would be included in the control group.

Table A5. Sample sizes for cohort studies^a

Incidence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	2.5	2.5
0.00001	1 472 091	2 082 958	4 710 686	21 983 178	28 264 016	7 851 105	2 355 325	1 221 276		
0.00005	294 411	416 580	942 108	4 396 481	5 652 548	1 570 142	471 036	244 238		
0.0001	147 201	208 283	471 036	2 198 144	2 826 115	785 022	235 500	122 108		
0.0005	29 433	41 645	94 178	439 474	564 968	156 925	47 071	24 404		
0.001	14 711	20 816	47 071	219 641	282 325	78 413	23 518	12 191		
0.005	2 935	4 152	9 385	43 774	56 210	15 604	4 675	2 421		
0.01	1 463	2 069	4 675	21 790	27 946	7 752	2 319	1 199		
0.05	285	402	906	4 204	5 334	1 471	435	222		
0.10	138	194	435	2 005	2 508	686	200	100		
0.15	89	125	278	1 273	1 566	425	121	59		
0.20	64	90	200	906	1 095	294	82	39		
0.25	49	69	152	686	812	215	58	27		
0.30	40	55	121	540	623	163	42	19		
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 A5. Sample sizes for cohort studies^a (continued)

Incidence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	785 104	565 273	436 166	294 411	157 946	106 615	45 666	16 672		
0.00005	157 008	113 044	87 224	58 875	31 583	21 318	9 129	3 330		
0.0001	78 496	56 515	43 606	29 433	15 788	10 656	4 562	1 663		
0.0005	15 686	11 292	8 712	5 879	3 152	2 126	908	329		
0.001	7 835	5 639	4 350	2 935	1 572	1 060	451	162		
0.005	1 554	1 117	861	579	309	207	86	28		
0.01	769	552	425	285	151	100	40	12		
0.05	141	100	76	49	24	15	3	—		
0.10	62	43	32	20	8	4	—	—		
0.15	36	24	17	10	—	—	—	—		
0.20	15	15	10	5	—	—	—	—		
0.25	10	9	6	—	—	—	—	—		
0.30	6	5	—	—	—	—	—	—		
0.35	—	—	—	—	—	—	—	—		
0.40	—	—	—	—	—	—	—	—		
0.45	—	—	—	—	—	—	—	—		
0.50	—	—	—	—	—	—	—	—		
0.55	—	—	—	—	—	—	—	—		
0.60	—	—	—	—	—	—	—	—		
0.65	—	—	—	—	—	—	—	—		
0.70	—	—	—	—	—	—	—	—		
0.75	—	—	—	—	—	—	—	—		
0.80	—	—	—	—	—	—	—	—		
0.85	—	—	—	—	—	—	—	—		
0.90	—	—	—	—	—	—	—	—		
0.95	—	—	—	—	—	—	—	—		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.20$ (power = 80%), control : exposed ratio = 1 : 1. The sample size listed is the number of subjects needed in the exposed group. An equivalent number would be included in the control group.

Table A6. Sample sizes for cohort studies^a

Incidence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	2.5	2.5
0.0001	1 190 356	1 663 432	3 680 447	16 792 779	20 878 641	5 726 194	1 683 582	859 799		
0.00005	238 065	332 677	736 066	3 358 436	4 175 543	1 145 183	336 697	171 948		
0.0001	119 028	166 332	368 018	1 679 143	2 087 655	572 556	168 336	85 967		
0.0005	23 799	33 257	73 580	335 708	417 346	114 455	33 648	17 182		
0.001	11 895	16 622	36 775	167 779	208 557	57 193	16 812	8 584		
0.005	2 372	3 315	7 332	33 436	41 526	11 382	3 343	1 705		
0.01	1 182	1 651	3 651	16 643	20 647	5 656	1 659	845		
0.05	230	321	707	3 208	3 944	1 075	312	157		
0.10	111	154	339	1 529	1 856	503	144	71		
0.15	71	99	216	969	1 160	312	88	43		
0.20	51	71	155	689	812	216	60	28		
0.25	39	54	118	522	603	159	43	20		
0.30	31	43	93	410	464	121	32	14		
0.35	26	35	76	330	365	93	23	10		
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 A6. Sample sizes for cohort studies^a (continued)

Incidence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	546 209	389 547	298 242	198 909	104 767	69 986	29 458	10 630		
0.00005	109 233	77 903	59 643	39 777	20 950	13 994	5 889	2 124		
0.0001	54 611	38 947	29 818	19 886	10 473	6 995	2 943	1 061		
0.0005	10 914	7 783	5 958	3 973	2 091	1 396	586	210		
0.001	5 452	3 887	2 975	1 983	1 043	696	292	104		
0.005	1 082	771	589	392	205	136	56	19		
0.01	536	381	291	193	100	66	26	8		
0.05	99	69	52	34	17	10	2	—		
0.10	44	30	23	14	6	3	—	—		
0.15	26	17	13	7	—	—	—	—		
0.20	17	11	8	4	—	—	—	—		
0.25	11	7	4	—	—	—	—	—		
0.30	7	4	—	—	—	—	—	—		
0.35	4	—	—	—	—	—	—	—		
0.40	—	—	—	—	—	—	—	—		
0.45	—	—	—	—	—	—	—	—		
0.50	—	—	—	—	—	—	—	—		
0.55	—	—	—	—	—	—	—	—		
0.60	—	—	—	—	—	—	—	—		
0.65	—	—	—	—	—	—	—	—		
0.70	—	—	—	—	—	—	—	—		
0.75	—	—	—	—	—	—	—	—		
0.80	—	—	—	—	—	—	—	—		
0.85	—	—	—	—	—	—	—	—		
0.90	—	—	—	—	—	—	—	—		
0.95	—	—	—	—	—	—	—	—		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.20$ (power = 80%), control : exposed ratio = 2 : 1. The sample size listed is the number of subjects needed in the exposed group. Double this number would be included in the control group.

Table A7. Sample sizes for cohort studies^a

Incidence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	4.0
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	—	—
0.00005	217 658	303 242	666 145	3 011 370	3 682 391	1 002 792	291 297	147 315	—	—
0.0001	108 825	151 615	333 059	1 505 617	1 841 094	501 366	145 638	73 651	—	—
0.0005	21 759	30 314	66 590	301 015	368 057	100 225	29 111	14 721	—	—
0.001	10 875	15 151	33 281	150 439	183 927	50 082	14 545	7 354	—	—
0.005	2 169	3 021	6 635	29 979	36 623	9 968	2 892	1 461	—	—
0.01	1 080	1 505	3 304	14 922	18 210	4 954	1 436	725	—	—
0.05	210	292	639	2 876	3 480	942	271	135	—	—
0.10	101	140	306	1 370	1 638	441	125	62	—	—
0.15	65	90	195	868	1 025	274	76	37	—	—
0.20	46	64	139	617	718	190	52	25	—	—
0.25	36	49	106	466	534	140	37	17	—	—
0.30	28	39	84	366	411	107	28	12	—	—
0.35	23	32	68	294	323	83	21	9	—	—
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 A7. Sample sizes for cohort studies^a (continued)

Incidence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	464 207	328 848	250 342	165 451	85 870	56 861	23 565	8 410		
0.00005	92 835	65 764	50 064	33 087	17 171	11 370	4 711	1 681		
0.0001	46 413	32 879	25 029	16 541	8 584	5 684	2 355	839		
0.0005	9 276	6 570	5 001	3 305	1 714	1 134	469	166		
0.001	4 634	3 282	2 498	1 650	855	566	233	82		
0.005	920	651	495	326	168	111	45	15		
0.01	456	322	245	161	83	54	21	6		
0.05	84	59	44	29	14	8	2	—		
0.10	38	26	19	12	5	2	—	—		
0.15	22	15	11	6	—	—	—	—		
0.20	14	9	6	3	—	—	—	—		
0.25	10	6	4	—	—	—	—	—		
0.30	6	3	—	—	—	—	—	—		
0.35	4	—	—	—	—	—	—	—		
0.40	—	—	—	—	—	—	—	—		
0.45	—	—	—	—	—	—	—	—		
0.50	—	—	—	—	—	—	—	—		
0.55	—	—	—	—	—	—	—	—		
0.60	—	—	—	—	—	—	—	—		
0.65	—	—	—	—	—	—	—	—		
0.70	—	—	—	—	—	—	—	—		
0.75	—	—	—	—	—	—	—	—		
0.80	—	—	—	—	—	—	—	—		
0.85	—	—	—	—	—	—	—	—		
0.90	—	—	—	—	—	—	—	—		
0.95	—	—	—	—	—	—	—	—		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.20$ (power = 80%), control : exposed ratio = 3 : 1. The sample size listed is the number of subjects needed in the exposed group. Triple this number would be included in the control group.

Table A8. Sample sizes for cohort studies^a

Incidence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	2.5	2.5
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	674 194	674 194
0.00005	206 915	288 054	630 802	2 837 520	3 435 570	931 374	268 406	134 830	134 830	134 830
0.0001	103 454	144 022	315 388	1 418 696	1 717 691	465 659	134 194	67 410	67 410	67 410
0.0005	20 685	28 795	63 057	283 636	343 387	93 087	26 824	13 473	13 473	13 473
0.001	10 338	14 392	31 515	141 754	171 599	46 516	13 402	6 731	6 731	6 731
0.005	2 061	2 870	6 282	28 248	34 169	9 259	2 665	1 338	1 338	1 338
0.01	1 027	1 429	3 128	14 059	16 990	4 601	1 323	663	663	663
0.05	199	277	605	2 709	3 247	876	250	124	124	124
0.10	96	133	289	1 290	1 529	410	115	57	57	57
0.15	61	85	184	817	957	255	71	34	34	34
0.20	44	61	132	581	670	177	48	23	23	23
0.25	34	47	100	439	499	130	35	16	16	16
0.30	27	37	79	344	384	99	26	11	11	11
0.35	22	30	64	277	302	77	19	8	8	8
0.40	18	25	53	226	241	60	14	5	5	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 A8. Sample sizes for cohort studies^a (continued)

Incidence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	422 454	297 814	225 764	148 182	76 019	49 975	20 438	7 223		
0.00005	84 485	59 558	45 149	29 633	15 201	9 993	4 086	1 443		
0.0001	42 238	29 776	22 572	14 815	7 599	4 995	2 042	721		
0.0005	8 442	5 950	4 510	2 960	1 518	997	407	143		
0.001	4 217	2 972	2 253	1 478	757	497	203	71		
0.005	837	590	446	292	149	98	39	13		
0.01	415	292	221	144	73	48	19	6		
0.05	77	53	40	26	12	8	2	—		
0.10	35	24	17	11	5	2	—	—		
0.15	20	14	10	6	—	—	—	—		
0.20	13	9	6	3	—	—	—	—		
0.25	9	5	3	—	—	—	—	—		
0.30	6	—	—	—	—	—	—	—		
0.35	3	—	—	—	—	—	—	—		
0.40	—	—	—	—	—	—	—	—		
0.45	—	—	—	—	—	—	—	—		
0.50	—	—	—	—	—	—	—	—		
0.55	—	—	—	—	—	—	—	—		
0.60	—	—	—	—	—	—	—	—		
0.65	—	—	—	—	—	—	—	—		
0.70	—	—	—	—	—	—	—	—		
0.75	—	—	—	—	—	—	—	—		
0.80	—	—	—	—	—	—	—	—		
0.85	—	—	—	—	—	—	—	—		
0.90	—	—	—	—	—	—	—	—		
0.95	—	—	—	—	—	—	—	—		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.20$ (power = 80%), control : exposed ratio = 4 : 1. The sample size listed is the number of subjects needed in the exposed group. Quadruple this number would be included in the control group.

Table A9. Sample sizes for case-control studies^a

Prevalence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5		
0.0001	1 970 728	2 788 519	6 306 363	29 429 793	37 838 497	10 510 715	3 153 225	1 635 011		
0.00005	394 143	557 705	1 261 292	5 886 130	7 568 072	2 102 264	630 690	327 029		
0.0001	197 070	278 853	630 659	2 943 172	3 784 269	1 051 207	315 373	163 532		
0.0005	39 412	55 772	126 151	588 806	757 227	210 362	63 120	32 734		
0.001	19 704	27 887	63 088	294 510	378 847	105 257	31 588	16 384		
0.005	3 939	5 579	12 638	59 074	76 145	21 173	6 363	3 304		
0.01	1 968	2 790	6 331	29 646	38 309	10 663	3 210	1 669		
0.05	391	560	1 288	6 111	8 059	2 261	690	363		
0.10	195	281	659	3 181	4 302	1 219	379	202		
0.15	129	189	451	2 215	3 072	879	278	150		
0.20	97	143	348	1 741	2 476	716	230	126		
0.25	77	116	287	1 465	2 137	624	203	113		
0.30	64	98	248	1 289	1 930	569	188	106		
0.35	56	86	222	1 174	1 802	536	180	103		
0.40	49	77	203	1 097	1 727	519	177	102		
0.45	44	70	191	1 048	1 694	513	178	104		
0.50	40	66	182	1 023	1 696	519	182	108		
0.55	38	62	178	1 019	1 732	535	191	114		
0.60	36	61	177	1 035	1 806	562	203	123		
0.65	35	60	180	1 077	1 927	605	222	135		
0.70	34	61	188	1 149	2 110	669	248	153		
0.75	35	64	203	1 268	2 390	764	287	178		
0.80	37	70	230	1 465	2 831	913	348	218		
0.85	43	82	278	1 811	3 591	1 168	451	285		
0.90	54	108	379	2 527	5 143	1 687	659	420		
0.95	93	190	690	4 717	9 851	3 257	1 288	828		

^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 A9. Sample sizes for case-control studies^a (continued)

Prevalence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	1 051 081	756 780	583 937	394 159	211 464	142 743	61 147	22 330		
0.00005	210 236	151 372	116 801	78 842	42 300	28 555	12 234	4 469		
0.0001	105 130	75 696	58 409	39 427	21 155	14 281	6 120	2 237		
0.0005	21 046	15 155	11 695	7 896	4 238	2 862	1 228	451		
0.001	10 535	7 587	5 856	3 954	2 124	1 435	617	228		
0.005	2 127	1 533	1 184	801	432	293	128	49		
0.01	1 076	777	601	407	221	150	67	27		
0.05	237	172	135	93	52	37	18	9		
0.10	133	98	77	54	32	23	13	8		
0.15	100	75	60	43	26	19	11	8		
0.20	85	64	52	37	23	18	11	8		
0.25	77	59	48	35	23	18	12	9		
0.30	73	56	46	34	23	18	13	10		
0.35	72	56	46	35	24	19	14	11		
0.40	72	56	47	36	25	20	15	12		
0.45	74	58	49	38	27	22	17	14		
0.50	77	61	52	40	29	24	19	16		
0.55	82	66	56	44	32	27	21	18		
0.60	89	72	61	49	36	31	25	21		
0.65	99	80	69	56	42	36	29	25		
0.70	113	92	79	64	49	43	35	31		
0.75	133	109	94	77	59	52	43	38		
0.80	164	135	117	97	75	66	55	49		
0.85	216	179	156	129	101	90	75	68		
0.90	320	266	233	195	154	137	116	105		
0.95	635	531	466	391	313	280	238	217		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.10$ (power = 90%), control : case ratio = 1 : 1. The sample size listed is the number of subjects needed in the case group. An equivalent number would be included in the control group.

Table A10. Sample sizes for case-control studies^a

Prevalence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	2.5	2.5
0.00001	1 529 065	2 153 652	4 825 672	22 280 178	28 149 758	7 764 749	2 302 966	1 183 610		
0.00005	305 811	430 731	965 148	4 456 162	5 630 233	1 553 041	460 628	236 743		
0.0001	152 904	215 366	482 583	2 228 160	2 815 293	776 578	230 335	118 385		
0.0005	30 578	43 073	96 531	445 759	563 340	155 407	46 101	23 698		
0.001	15 288	21 537	48 274	222 959	281 846	77 761	23 072	11 862		
0.005	3 055	4 308	9 669	44 719	56 653	15 644	4 649	2 393		
0.01	1 526	2 154	4 843	22 440	28 505	7 880	2 346	1 210		
0.05	303	431	984	4 623	6 001	1 674	506	264		
0.10	150	216	503	2 405	3 207	904	279	148		
0.15	100	145	343	1 673	2 292	653	205	111		
0.20	74	110	265	1 313	1 849	533	170	93		
0.25	59	89	218	1 104	1 597	465	151	84		
0.30	49	75	188	971	1 443	425	140	79		
0.35	42	65	168	883	1 349	401	135	77		
0.40	37	58	154	825	1 294	388	133	77		
0.45	33	53	144	788	1 270	385	133	78		
0.50	31	50	137	768	1 272	389	137	81		
0.55	28	47	133	764	1 301	402	144	86		
0.60	27	45	133	775	1 357	423	154	93		
0.65	26	45	135	805	1 449	456	168	103		
0.70	26	45	140	859	1 588	505	188	116		
0.75	26	47	151	947	1 799	577	218	136		
0.80	28	51	170	1 092	2 133	690	265	166		
0.85	31	60	205	1 349	2 708	884	343	218		
0.90	39	78	279	1 880	3 881	1 278	503	322		
0.95	66	137	506	3 505	7 438	2 472	984	635		

^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 A10. Sample sizes for case-control studies^a (continued)

Prevalence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	755 564	540 911	415 405	278 348	147 639	99 012	41 948	15 205		
0.00005	151 128	108 194	83 091	55 678	29 534	19 807	8 393	3 044		
0.0001	75 573	54 105	41 552	27 844	14 770	9 906	4 199	1 524		
0.0005	15 130	10 833	8 321	5 577	2 960	1 986	843	307		
0.001	7 574	5 424	4 167	2 793	1 483	996	424	155		
0.005	1 530	1 097	844	567	302	204	88	34		
0.01	775	556	428	289	155	105	46	19		
0.05	171	124	97	66	37	26	13	7		
0.10	97	71	56	39	23	17	9	6		
0.15	74	55	44	31	19	14	8	6		
0.20	63	47	38	28	17	13	8	6		
0.25	57	44	35	26	17	13	9	6		
0.30	55	42	34	26	17	14	9	7		
0.35	54	42	34	26	18	14	10	8		
0.40	54	42	35	27	19	15	11	9		
0.45	56	44	37	28	20	17	13	10		
0.50	58	46	39	31	22	19	14	12		
0.55	62	50	42	33	24	21	16	14		
0.60	68	55	47	37	28	24	19	16		
0.65	76	61	52	42	32	28	22	19		
0.70	86	70	61	49	38	33	27	23		
0.75	102	84	72	59	46	40	33	29		
0.80	125	104	90	74	58	51	42	38		
0.85	165	137	120	100	78	70	58	53		
0.90	246	205	180	150	119	107	90	82		
0.95	489	410	360	303	243	218	186	169		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.10$ (power = 90%), control : case ratio = 2 : 1. The sample size listed is the number of subjects needed in the case group. Double this number would be included in the control group.

Table A11. Sample sizes for case-control studies^a

Prevalence 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	4.0
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		
0.00005	273 893	386 172	864 545	3 977 907	4 983 044	1 368 844	402 996	205 830		
0.0001	136 945	193 086	432 280	1 989 023	2 491 679	684 473	201 517	102 927		
0.0005	27 387	38 617	86 468	397 917	498 587	136 977	40 334	20 604		
0.001	13 692	19 309	43 242	199 028	249 451	68 540	20 186	10 314		
0.005	2 736	3 862	8 661	39 918	50 143	13 790	4 068	2 082		
0.01	1 367	1 931	4 338	20 030	25 231	6 947	2 054	1 053		
0.05	271	387	881	4 125	5 313	1 477	444	231		
0.10	134	194	450	2 145	2 841	799	245	129		
0.15	89	130	307	1 491	2 031	577	180	97		
0.20	66	98	236	1 171	1 639	471	150	82		
0.25	53	79	195	984	1 417	412	133	74		
0.30	44	67	168	865	1 281	376	124	70		
0.35	38	58	150	786	1 197	355	119	68		
0.40	33	52	137	734	1 149	345	118	68		
0.45	30	47	128	700	1 128	342	119	69		
0.50	27	44	122	682	1 131	346	122	72		
0.55	25	42	119	679	1 156	357	128	76		
0.60	24	40	118	689	1 207	377	137	83		
0.65	23	40	119	715	1 289	406	150	91		
0.70	23	40	124	762	1 414	450	168	104		
0.75	23	42	133	839	1 602	515	195	121		
0.80	24	45	150	968	1 900	616	236	149		
0.85	27	52	180	1 194	2 413	789	307	195		
0.90	34	68	245	1 664	3 459	1 142	450	288		
0.95	57	119	444	3 100	6 632	2 208	881	569		

^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 A11. Sample sizes for case-control studies^a (continued)

Prevalence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	653 448	465 720	356 295	237 271	124 583	83 040	34 800	12 517		
0.00005	130 703	93 155	71 268	47 461	24 922	16 612	6 963	2 506		
0.0001	65 360	46 584	35 640	23 735	12 464	8 309	3 483	1 254		
0.0005	13 086	9 328	7 137	4 754	2 498	1 666	700	253		
0.001	6 551	4 671	3 574	2 382	1 252	836	352	128		
0.005	1 324	945	724	484	256	171	73	28		
0.01	671	480	368	246	131	88	39	16		
0.05	149	108	84	57	32	22	11	6		
0.10	85	62	49	34	20	14	8	5		
0.15	64	48	38	27	16	12	7	5		
0.20	55	41	33	24	15	12	7	5		
0.25	50	38	31	23	15	12	8	6		
0.30	48	37	30	23	15	12	8	6		
0.35	47	37	30	23	16	13	9	7		
0.40	48	37	31	24	16	14	10	8		
0.45	49	39	32	25	18	15	11	9		
0.50	52	41	35	27	19	16	12	10		
0.55	55	44	38	30	22	18	14	12		
0.60	60	49	41	33	25	21	17	14		
0.65	67	55	47	38	28	24	20	17		
0.70	77	63	54	44	33	29	24	21		
0.75	91	75	65	53	41	36	29	26		
0.80	112	93	80	66	52	45	38	34		
0.85	148	123	107	89	70	62	52	46		
0.90	220	184	161	134	107	95	80	72		
0.95	438	367	323	271	217	194	165	150		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.10$ (power = 90%), control : case ratio = 3 : 1. The sample size listed is the number of subjects needed in the case group. Triple this number would be included in the control group.

Table A12. Sample sizes for case-control studies^a

Prevalence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5		
0.00001	1 285 573	1 815 890	4 068 256	18 690 963	23 294 197	6 381 647	1 869 301	950 501		
0.00005	257 112	363 178	813 662	3 738 297	4 659 075	1 276 406	373 889	190 118		
0.0001	128 555	181 589	406 838	1 869 214	2 329 685	638 251	186 963	95 070		
0.0005	25 709	36 318	81 379	373 947	466 173	127 727	37 422	19 032		
0.001	12 853	18 159	40 697	187 039	233 234	63 912	18 729	9 527		
0.005	2 568	3 632	8 151	37 513	46 884	12 860	3 775	1 923		
0.01	1 283	1 816	4 082	18 823	23 592	6 479	1 906	973		
0.05	255	363	829	3 876	4 969	1 378	412	214		
0.10	126	182	423	2 015	2 658	746	228	120		
0.15	83	122	289	1 401	1 901	539	168	90		
0.20	62	92	222	1 099	1 534	440	140	76		
0.25	50	74	183	923	1 326	385	125	69		
0.30	41	63	158	812	1 200	352	116	65		
0.35	35	55	140	738	1 122	333	111	63		
0.40	31	49	128	688	1 077	323	110	63		
0.45	28	44	120	657	1 058	320	111	65		
0.50	25	41	114	640	1 060	324	114	67		
0.55	23	39	111	636	1 084	335	120	72		
0.60	22	38	110	645	1 132	354	128	78		
0.65	21	37	111	669	1 209	381	140	86		
0.70	21	37	116	713	1 326	422	158	97		
0.75	21	39	125	786	1 504	483	183	114		
0.80	22	42	140	905	1 784	579	222	140		
0.85	25	48	168	1 117	2 266	742	289	183		
0.90	31	63	228	1 556	3 248	1 073	423	271		
0.95	52	110	412	2 897	6 229	2 076	829	536		

^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 A12. Sample sizes for case-control studies^a (continued)

Prevalence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	601 245	427 084	325 786	215 910	112 440	74 563	30 952	11 054		
0.00005	120 262	85 427	65 166	43 189	22 493	14 916	6 193	2 213		
0.0001	60 139	42 720	32 588	21 599	11 249	7 461	3 098	1 108		
0.0005	12 041	8 554	6 526	4 326	2 255	1 496	622	224		
0.001	6 028	4 284	3 269	2 167	1 130	750	313	113		
0.005	1 219	867	662	440	231	154	65	25		
0.01	618	440	337	224	118	79	34	14		
0.05	137	99	77	52	29	20	10	5		
0.10	78	57	45	31	18	13	7	4		
0.15	60	44	35	25	15	11	7	4		
0.20	51	38	31	22	14	11	7	5		
0.25	47	36	29	21	14	11	7	5		
0.30	45	34	28	21	14	11	7	6		
0.35	44	34	28	21	14	12	8	6		
0.40	45	35	29	22	15	13	9	7		
0.45	46	36	30	23	17	14	10	8		
0.50	48	38	32	25	18	15	11	10		
0.55	52	41	35	28	20	17	13	11		
0.60	57	46	39	31	23	20	15	13		
0.65	63	51	44	35	26	23	18	16		
0.70	72	59	51	41	31	27	22	19		
0.75	85	70	61	50	38	33	27	24		
0.80	105	87	75	62	48	42	35	31		
0.85	139	115	101	83	65	58	48	43		
0.90	207	173	151	126	100	89	75	67		
0.95	412	345	303	255	203	182	154	139		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.10$ (power = 90%), control : case ratio = 4 : 1. The sample size listed is the number of subjects needed in the case group. Quadruple this number would be included in the control group.

Table A13. Sample sizes for case-control studies^a

Prevalence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	4.0
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		
0.00005	294 418	416 596	942 163	4 396 835	5 653 216	1 570 354	471 115	244 286		
0.0001	147 208	208 299	471 091	2 198 497	2 826 782	785 234	235 579	122 156		
0.0005	29 440	41 661	94 233	439 828	565 636	157 137	47 150	24 452		
0.001	14 719	20 831	47 126	219 994	282 992	78 625	23 596	12 239		
0.005	2 943	4 168	9 441	44 128	56 879	15 816	4 753	2 469		
0.01	1 470	2 085	4 730	22 145	28 617	7 966	2 398	1 248		
0.05	293	419	962	4 566	6 020	1 690	516	272		
0.10	146	211	493	2 377	3 214	911	283	151		
0.15	97	142	337	1 655	2 295	657	208	113		
0.20	73	107	260	1 301	1 850	535	172	95		
0.25	58	87	215	1 095	1 597	466	152	85		
0.30	49	74	186	964	1 442	425	141	80		
0.35	42	65	166	877	1 346	401	135	77		
0.40	37	58	152	820	1 291	388	133	77		
0.45	33	53	143	784	1 266	384	133	78		
0.50	31	50	137	765	1 267	388	137	81		
0.55	29	47	133	761	1 294	400	143	85		
0.60	27	46	133	774	1 350	421	152	92		
0.65	26	45	135	805	1 440	453	166	101		
0.70	26	46	141	859	1 577	500	186	115		
0.75	27	48	152	948	1 785	571	215	134		
0.80	28	53	172	1 095	2 115	682	260	163		
0.85	32	62	208	1 353	2 683	873	337	213		
0.90	41	81	283	1 888	3 842	1 260	493	314		
0.95	70	142	516	3 524	7 359	2 433	962	619		

^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 A13. Sample sizes for case--control studies^a (continued)

Prevalence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	785	565	436	294	157	106	45	16		
0.00005	157	113	87	58	31	21	9	3		
0.0001	78	56	43	29	15	10	4	1		
0.0005	15	11	8	5	3	2	1	0		
0.001	7	5	4	2	1	1	0	0		
0.005	1	1	0	0	0	0	0	0		
0.01	804	581	449	305	165	113	50	20		
0.05	177	129	101	70	39	28	14	7		
0.10	100	74	58	41	24	18	10	6		
0.15	75	56	45	32	20	15	9	6		
0.20	64	48	39	28	18	14	9	6		
0.25	58	44	36	27	17	14	9	7		
0.30	55	42	35	26	18	14	10	8		
0.35	54	42	35	26	18	15	11	9		
0.40	54	42	35	27	19	16	12	10		
0.45	56	44	37	29	20	17	13	11		
0.50	58	46	39	31	22	19	15	12		
0.55	62	50	42	33	25	21	16	14		
0.60	67	54	46	37	28	24	19	16		
0.65	75	61	52	42	32	27	22	19		
0.70	85	69	60	49	37	32	26	23		
0.75	100	82	71	58	45	39	33	29		
0.80	123	101	88	73	57	50	42	37		
0.85	162	134	117	97	76	68	57	51		
0.90	240	200	175	146	116	103	87	79		
0.95	475	397	349	293	234	210	179	162		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.20$ (power = 80%), control : case ratio = 1 : 1. The sample size listed is the number of subjects needed in the case group. An equivalent number would be included in the control group.

Table A14. Sample sizes for case-control studies^a

Prevalence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	2.5	2.5
0.00001	1 190 363	1 663 444	3 680 489	16 793 046	20 879 138	5 726 351	1 683 639	859 834		
0.00005	238 071	332 689	736 108	3 358 703	4 176 039	1 145 339	336 754	171 983		
0.0001	119 034	166 344	368 060	1 679 410	2 088 152	572 713	168 393	86 001		
0.0005	23 805	33 269	73 622	335 976	417 842	114 612	33 705	17 216		
0.001	11 901	16 635	36 817	168 047	209 054	57 349	16 869	8 618		
0.005	2 378	3 327	7 374	33 704	42 024	11 540	3 400	1 740		
0.01	1 188	1 664	3 693	16 911	21 146	5 814	1 717	880		
0.05	236	333	750	3 482	4 455	1 237	371	193		
0.10	117	167	383	1 810	2 383	669	205	109		
0.15	77	112	261	1 258	1 704	484	152	82		
0.20	58	84	201	987	1 376	396	126	69		
0.25	46	68	166	829	1 190	346	112	62		
0.30	38	57	143	729	1 076	316	105	59		
0.35	33	50	127	662	1 006	299	101	58		
0.40	29	45	116	618	966	290	99	58		
0.45	26	41	108	590	949	288	100	59		
0.50	24	38	103	574	951	292	103	61		
0.55	22	36	100	571	973	301	108	65		
0.60	21	34	99	579	1 016	318	116	70		
0.65	20	34	101	601	1 085	343	127	78		
0.70	20	34	105	640	1 190	380	143	89		
0.75	20	35	112	705	1 350	435	166	104		
0.80	21	38	126	812	1 601	520	201	127		
0.85	23	44	152	1 002	2 034	667	261	167		
0.90	29	58	205	1 395	2 916	965	383	246		
0.95	48	100	371	2 598	5 592	1 868	750	487		

^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 A14. Sample sizes for case-control studies^a (continued)

Prevalence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	546 235	389 568	298 260	198 923	104 777	69 995	29 465	10 635		
0.00005	109 259	77 923	59 660	39 791	20 960	14 003	5 896	2 129		
0.0001	54 637	38 967	29 835	19 899	10 483	7 004	2 950	1 066		
0.0005	10 939	7 803	5 975	3 986	2 101	1 405	593	216		
0.001	5 477	3 907	2 993	1 997	1 053	705	298	109		
0.005	1 107	791	607	406	215	145	63	24		
0.01	561	402	308	207	111	75	33	14		
0.05	125	91	70	48	27	19	10	5		
0.10	71	53	41	29	17	13	7	5		
0.15	54	41	32	23	14	11	7	5		
0.20	47	35	28	21	13	10	7	5		
0.25	43	33	27	20	13	10	7	5		
0.30	41	32	26	20	13	11	7	6		
0.35	40	31	26	20	14	11	8	7		
0.40	41	32	27	21	15	12	9	7		
0.45	42	33	28	22	16	13	10	8		
0.50	44	35	30	24	17	15	11	10		
0.55	47	38	32	26	19	16	13	11		
0.60	52	42	36	29	22	19	15	13		
0.65	58	47	40	33	25	22	18	16		
0.70	66	54	47	38	29	26	21	19		
0.75	78	64	56	46	36	32	26	23		
0.80	96	80	70	58	45	40	34	30		
0.85	127	106	93	77	61	55	46	42		
0.90	189	158	139	117	94	84	71	65		
0.95	376	316	279	236	190	171	147	134		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.20$ (power = 80%), control: case ratio = 2:1. The sample size listed is the number of subjects needed in the case group. Double this number would be included in the control group.

Table A15. Sample sizes for case-control studies^a

Prevalence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	3.0	4.0
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		
0.00005	217 664	303 253	666 182	3 011 608	3 682 831	1 002 930	291 347	147 345		
0.0001	108 831	151 626	333 096	1 505 856	1 841 534	501 504	145 688	73 681		
0.0005	21 764	30 325	66 628	301 253	368 496	100 363	29 161	14 751		
0.001	10 881	15 162	33 319	150 678	184 367	50 220	14 595	7 384		
0.005	2 174	3 032	6 672	30 218	37 064	10 107	2 943	1 491		
0.01	1 086	1 516	3 342	15 161	18 652	5 093	1 486	755		
0.05	215	303	678	3 120	3 932	1 085	323	167		
0.10	107	152	345	1 620	2 105	588	179	94		
0.15	70	101	235	1 125	1 507	426	132	71		
0.20	52	76	181	882	1 218	349	111	60		
0.25	42	62	149	741	1 053	305	99	55		
0.30	35	52	128	650	954	280	92	52		
0.35	30	45	114	590	892	265	89	51		
0.40	26	40	104	550	857	257	88	51		
0.45	23	36	97	525	843	256	89	52		
0.50	21	34	92	511	846	259	92	55		
0.55	19	32	89	507	866	268	97	58		
0.60	18	30	88	514	905	283	104	63		
0.65	18	30	89	533	967	306	114	70		
0.70	17	30	92	567	1 061	339	128	80		
0.75	17	31	99	624	1 204	389	149	93		
0.80	18	33	111	718	1 429	466	181	115		
0.85	20	38	132	884	1 817	598	235	151		
0.90	25	50	179	1 230	2 607	867	345	223		
0.95	41	85	323	2 288	5 002	1 678	678	442		

^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 A15. Sample sizes for case-control studies^a (continued)

Prevalence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	464 229	328 865	250 357	165 463	85 879	56 868	23 570	8 415		
0.00005	92 856	65 782	50 079	33 098	17 180	11 377	4 717	1 685		
0.0001	46 435	32 896	25 044	16 553	8 592	5 691	2 360	844		
0.0005	9 298	6 588	5 016	3 316	1 723	1 141	474	171		
0.001	4 655	3 299	2 513	1 662	864	573	239	87		
0.005	942	668	510	338	177	118	50	19		
0.01	478	340	259	173	91	61	27	11		
0.05	107	77	60	41	23	16	8	4		
0.10	62	45	35	25	15	11	6	4		
0.15	47	35	28	20	12	9	6	4		
0.20	41	31	25	18	11	9	6	4		
0.25	37	29	23	17	11	9	6	5		
0.30	36	28	23	17	12	9	7	5		
0.35	36	28	23	18	12	10	7	6		
0.40	36	28	24	18	13	11	8	7		
0.45	37	30	25	19	14	12	9	7		
0.50	39	32	27	21	15	13	10	9		
0.55	42	34	29	23	17	15	12	10		
0.60	46	38	32	26	19	17	13	12		
0.65	52	42	36	30	23	20	16	14		
0.70	60	49	42	35	27	23	19	17		
0.75	70	58	51	42	32	29	24	21		
0.80	87	72	63	52	41	37	31	28		
0.85	115	96	84	70	56	50	42	38		
0.90	172	144	127	107	85	77	65	59		
0.95	342	288	255	215	174	157	134	123		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.20$ (power = 80%), control : case ratio = 3 : 1. The sample size listed is the number of subjects needed in the case group. Triple this number would be included in the control group.

Table A16. Sample sizes for case-control studies^a

Prevalence in control group	Relative risk to be detected									
	0.2	0.3	0.5	0.75	1.25	1.5	2.0	2.5	2.5	2.5
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		
0.00005	206 920	288 065	630 838	2 837 745	3 435 981	931 503	268 452	134 858		
0.0001	103 459	144 032	315 424	1 418 920	1 718 102	465 788	134 240	67 438		
0.0005	20 690	28 806	63 092	283 861	343 799	93 216	26 870	13 501		
0.001	10 344	14 403	31 551	141 978	172 011	46 645	13 449	6 759		
0.005	2 067	2 880	6 318	28 473	34 581	9 388	2 712	1 366		
0.01	1 032	1 440	3 164	14 285	17 404	4 731	1 370	691		
0.05	205	288	641	2 938	3 670	1 009	298	153		
0.10	101	144	327	1 525	1 966	547	166	87		
0.15	67	96	222	1 059	1 408	397	123	66		
0.20	50	72	171	830	1 138	325	103	56		
0.25	39	58	140	696	985	285	92	51		
0.30	33	49	121	611	892	261	86	48		
0.35	28	42	107	554	836	248	83	47		
0.40	24	38	97	517	803	241	82	48		
0.45	22	34	91	492	790	240	83	49		
0.50	20	32	86	479	793	243	86	51		
0.55	18	30	83	475	812	252	91	55		
0.60	17	28	82	481	849	266	97	59		
0.65	16	28	83	498	908	288	107	66		
0.70	16	28	86	530	997	319	121	75		
0.75	16	29	92	583	1 131	366	140	88		
0.80	17	31	103	670	1 343	439	171	108		
0.85	18	35	123	826	1 708	564	222	143		
0.90	23	45	166	1 148	2 452	817	327	211		
0.95	37	78	298	2 133	4 707	1 583	641	419		

^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 A16. Sample sizes for case-control studies^a (continued)

Prevalence in control group	Relative risk to be detected									
	3.0	3.5	4.0	5.0	7.5	10.0	20.0	50.0		
0.00001	422 474	297 830	225 778	148 193	76 026	49 982	20 443	7 227		
0.00005	84 505	59 574	45 162	29 644	15 209	9 999	4 091	1 447		
0.0001	42 259	29 792	22 585	14 825	7 607	5 002	2 047	725		
0.0005	8 462	5 966	4 524	2 970	1 525	1 003	412	147		
0.001	4 237	2 988	2 266	1 489	765	504	207	75		
0.005	858	606	460	303	157	104	44	17		
0.01	435	308	234	155	81	54	23	10		
0.05	98	70	54	37	20	14	7	4		
0.10	57	41	32	23	13	10	5	3		
0.15	43	32	26	18	11	8	5	4		
0.20	38	28	23	17	10	8	5	4		
0.25	35	26	21	16	10	8	6	4		
0.30	33	26	21	16	11	9	6	5		
0.35	33	26	21	16	11	9	7	5		
0.40	34	26	22	17	12	10	7	6		
0.45	35	28	23	18	13	11	8	7		
0.50	37	30	25	20	14	12	9	8		
0.55	40	32	27	22	16	14	11	9		
0.60	44	35	30	24	18	16	13	11		
0.65	49	40	34	28	21	18	15	13		
0.70	56	46	40	33	25	22	18	16		
0.75	67	55	48	39	31	27	22	20		
0.80	82	68	60	50	39	35	29	26		
0.85	109	91	80	67	53	47	40	36		
0.90	163	137	120	101	81	73	62	56		
0.95	325	273	242	205	165	149	127	116		

^a $\alpha = 0.05$ (two tailed), $\beta = 0.20$ (power = 80%), control : case ratio = 4 : 1. The sample size listed is the number of subjects needed in the case group. Quadruple this number would be included in the control group.

Table A17. Tabular values of 95 percent confidence limit factors for estimates of a Poisson distributed variable^a

Observed number on which estimate is based (n)	Observed number on which estimate is based (n)		Observed number on which estimate is based (n)	
	Lower limit factor (L)	Upper limit factor (U)	Lower limit factor (L)	Upper limit factor (U)
1	0.0253	5.57	0.619	1.53
2	0.121	3.61	0.627	1.51
3	0.206	2.92	0.634	1.50
4	0.272	2.56	0.641	1.49
5	0.324	2.33	0.647	1.48
6	0.367	2.18	0.653	1.47
7	0.401	2.06	0.659	1.46
8	0.431	1.97	0.665	1.45
9	0.458	1.90	0.670	1.44
10	0.480	1.84	0.675	1.43
11	0.499	1.79	0.697	1.39
12	0.517	1.75	0.714	1.36
13	0.532	1.71	0.729	1.34
14	0.546	1.68	0.742	1.32
15	0.560	1.65	0.770	1.30
16	0.572	1.62	0.785	1.27
17	0.583	1.60	0.798	1.25
18	0.593	1.58	0.809	1.24
19	0.602	1.56	0.818	1.22
20	0.611	1.54		

^aHaenszel 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.

Appendix B

Glossary

The *accuracy* of a measurement is the degree to which the measurement approximates the truth.

An *adverse drug event* or *adverse drug experience* is an adverse outcome that occurs during or following clinical use of a drug.

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, including the following: an adverse experience occurring in the course of the use of the product in professional practice; an adverse experience occurring from overdose of the product whether accidental or intentional; an adverse experience occurring from abuse of the product; an adverse experience occurring from withdrawal of the product; and any failure of expected pharmacological action.

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 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.

Analytic studies are studies with control groups, namely case–control studies, cohort studies, and randomized clinical trials.

Anticipated beneficial effects of drugs are desirable effects that are known 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 a population, that is an association, to a causal statement about biological theory.

Case–control studies are studies that compare cases with a disease to controls without the disease, 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, usually adverse, outcome.

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 occurs when a drug that is claimed to be very safe may first be tried on patients who do not tolerate the previous drug products. It is a type of selection bias.

Drug *clearance* is the proportion of the “apparent” volume of distribution that is cleared of drug in 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 intervals are a range of values within which the true population value probably lies.

Confidentiality is the right to limit the transfer 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 an apparent association between the risk factor and the outcome or mask a real one.

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-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 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.

Drug utilization evaluation studies are studies 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 studies that assess the appropriateness of drug use. They are designed to detect and quantify any drug use problems.

Drug utilization studies are descriptive studies that quantify the use of a drug. The objective of the study is to quantify the present state, the developmental trends, and the time course of drug usage at various levels of the health care system, whether national, regional, local, or institutional.

Ecological studies examine trends in disease events over time or across different geographic locations and correlate them with trends in a putative exposure, such as rates of drug utilization. The unit of observation is a subgroup of a population, rather than individuals.

Effect modification occurs when the magnitude of effect of a drug in causing an outcome differs according to the level of a variable other than the drug or the outcome.

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 judgement 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 incurred regardless of the volume of activity.

Generic quality of life instruments cover (or at least aim to cover) the complete spectrum of function, disability, and distress of the patient, and are applicable to a variety of populations.

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 that, from the patient's perspective, represents the end result of all the physiological, psychological, and social influences of 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, sexual function within physical and functional domains.

A *human research subject* is “a living individual, about whom an investigator (whether professional or student) conducting research obtains either: 1) data through intervention or interaction with the individual, or 2) identifiable private information.”

Hypothesis-generating studies are studies that give rise to new questions about drug effects to be explored further in subsequent studies.

Hypothesis-strengthening studies are studies that reinforce, although they 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 that 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 exposure or outcome.

Intangible costs are those of pain, suffering, and grief.

Interaction, see effect modification.

Interrupted time-series designs include multiple observations (often ten or more) of study populations before and after an intervention.

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 application of the methods of molecular biology in a pharmacoepidemiology study.

An *N-of-1 RCT* is a randomized controlled trial in an individual patient.

Nondifferential misclassification occurs when the misclassification of one variable does not vary by the level of another variable. The measure of association is usually biased toward the null.

Nonexperimental studies are studies in which the investigator does not control the therapy, but observes and evaluates the results of ongoing medical care. The study designs that are used are those that do not involve random allocation, 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. The study designs that are used are those 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 that has already been exposed to a treatment.

An *opportunity cost* is the value of a resource's next best use, a use no longer possible once the resource has been used.

A *p-value* is the probability that a difference as large as the one observed 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.

Pharmacogenetics is the study of genetic determinants of responses to drugs.

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.

Postmarketing surveillance is the study of drug use and drug effects after release onto the market. This term is sometimes used synonymously with “pharmacoepidemiology,” but the latter can be relevant to premarketing studies, as well. Conversely, the term “postmarketing surveillance” is sometimes felt to apply to only those studies conducted after drug marketing which 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.

The *power* of a study is the probability of detecting a difference in the study if one really exists.

Precision is the degree of absence of random error.

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.

The *prevalence rate* 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, 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.

Privacy, in the setting of research, refers to each individual's right to be free from unwanted inspection of, or access to, personal information.

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.

Protopathic bias is interpreting as a result of an exposure a variable that is in fact its determinant.

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 controls the therapy that is to be received by each participant and uses that control to allocate patients among the study groups randomly.

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.

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 is any activity designed to “develop or contribute to generalizable knowledge.”

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.

Risk is the probability that something will happen.

A judgement about *safety* is a personal and/or social judgement 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, that 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 judgement 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 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 as not having it.

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.

Within pharmacokinetics, *steady state* is the situation when the amount of drug being administered is equal to 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 which are aberrant effects of the drug. They tend to be uncommon, not dose-related, and unpredictable.

A *type I error* is concluding there is an association when in fact one does not exist.

A *type II error* is concluding there is no association when in fact one does exist.

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.

An *unexpected adverse experience* means any adverse experience 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.

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 elements of a utility instrument are first, that it is preference based, and second, that scores are tied to death as an outcome.

Validity is the degree to which an assessment 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.