



Milbank Memorial Fund

Evaluating Health Services: A Reporter Covers the Science of Research Synthesis

by Ray Moynihan

Evaluating Health Services: A Reporter Covers the Science of Research Synthesis

by Ray Moynihan

©2004 Milbank Memorial Fund

Milbank Memorial Fund
645 Madison Avenue
New York, NY 10022

All rights reserved. No part of this publication may be reproduced, stored in any retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior permission.

The Milbank Memorial Fund is an endowed operating foundation that engages in nonpartisan analysis, study, research, and communication on significant issues in health policy. In the Fund's own publications, in reports or books it publishes with other organizations, and in articles it commissions for publication by other organizations, the Fund endeavors to maintain the highest standards for accuracy and fairness. Statements by individual authors, however, do not necessarily reflect opinions or factual determinations of the Fund.

Printed in the United States of America.

ISBN 1-887748-56-3

TABLE OF CONTENTS

Foreword	v
Acknowledgments	vi
Executive Summary	1
A Note About Nomenclature	2
Introduction to Systematic Reviews	3
The History of Systematic Reviews	6
Why Do Systematic Reviews?	10
How to Do a Systematic Review	14
The Uses of Systematic Reviews	21
The Limitations of Systematic Reviews	27
The Future of Systematic Reviews	34
Notes	36
References	42
The Author	51

FOREWORD

Systematic reviews, based on the science of research synthesis, are an increasingly influential source of useful information about the effectiveness of interventions in health care and other areas of public concern. Scientists in many countries are using the methods of research synthesis to prepare rigorous systematic reviews of evidence from published and unpublished studies. These reviews are informing the judgments of many clinicians and policymakers about which interventions are likely to be most appropriate for particular patients and populations.

The Milbank Memorial Fund asked Ray Moynihan, a distinguished journalist, to report on these recent developments in the use of research to inform practice and policy. Because many people concerned about health affairs have not followed recent developments in the methods and uses of research synthesis, the Fund asked Moynihan to describe who is doing systematic reviews, how the most reliable reviews are being prepared, and how they are being used. The Fund also asked him to report on how scientists and policymakers assess the strengths, weaknesses, and promise of the methodology for preparing reviews.

Moynihan has been covering the story of research synthesis and the move toward evidence-based medicine for many years, initially in Australia and then internationally. For this report, he interviewed pioneers in the science of research synthesis, clinicians and policymakers who use systematic reviews, and journalists and other persons who communicate about the findings of scientists to various audiences.

The Milbank Memorial Fund is an endowed operating foundation that engages in nonpartisan analysis, study, research, and communication on significant issues in health policy. Since 1905 the Fund has worked to improve and maintain health by encouraging persons who make and implement health policy to use the best available evidence. The Fund convenes meetings of leaders in the public and private sectors and publishes reports, articles, and books.

The Fund sought comments on this report in draft from many people in science, clinical practice, policy, and communications. These reviewers are listed in the Acknowledgments.

Daniel M. Fox
President

Samuel L. Milbank
Chairman

ACKNOWLEDGMENTS

The following persons were interviewed for this report and/or reviewed it in draft. They are listed in the positions they held at the time of their participation.

David Armstrong, Reader in Sociology as applied to Medicine, Department of General Practice, The Guy's, King's and St. Thomas' School of Medicine, London, U.K.; Naomi Aronson, Director, Evidence-Based Practice Center, Blue Cross Blue Shield Association, Chicago; Grant Bagley, Partner, Arnold and Porter, LLP, Washington, D.C.; John R. Ball, Executive Vice President, American Society for Clinical Pathology, Chicago; Lisa A. Bero, Associate Professor, Department of Clinical Pharmacy, Institute for Health Policy Studies, University of California at San Francisco; Edgar Black, Vice President and Chief Medical Officer, Blue Cross Blue Shield, Rochester, N.Y.; Nick Black, Professor, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine; Robert H. Brook, Professor, Medicine and Health Services, University of California at Los Angeles; Lanier Cansler, Deputy Secretary, North Carolina Department of Health and Human Services, Raleigh; Iain Chalmers, Editor, James Lind Library, Oxford, U.K.; Carolyn M. Clancy, Director, Agency for Healthcare Research and Quality, Rockville, Md.; Mike Clarke, Director, Cochrane Center, London, U.K.; Helen Darling, President, National Business Group on Health, Washington, D.C.; Robert M. Day, Medicaid Director, Kansas Department of Social and Rehabilitation Services, Topeka; Marilyn Desmond, Assistant Director, Policy and Program Guidance Division, Office of Medicaid Management, New York Department of Health, Albany; Kay Dickersin, Associate Professor, Department of Community Health, Brown University School of Medicine, Providence, R.I.; Jack C. Ebeler, President and Chief Executive Officer, Alliance of Community Health Plans, Washington, D.C.; David M. Eddy, Senior Advisor for Health Policy and Management, Kaiser Permanente, Oakland, Calif.; Nancy Ellison, Deputy Secretary, Washington Department of Health, Olympia; Karen M. Facey, Interim Director, Health Technology Assessment Board for Scotland, NHS, Glasgow; Christina Farup, Worldwide Vice President, Health Economics and Reimbursement, Ethicon, Somerville, N.J.; John Folkemer, Executive Director, Office of Planning and Finance, Maryland Department of Health and Mental Hygiene, Baltimore; Mark Gibson, Deputy Director, Center for Evidence-Based Policy, Oregon Health and Science University, Portland, and Program Officer, Milbank Memorial Fund; Bruce Goldberg, Administrator, Oregon Office of Health Policy and Research, Salem; Howard Goldman, Professor of Psychiatry, University of Maryland School of Medicine, Potomac; Richard N. Gottfried, Chair, Health Committee, New York State Assembly; Robert Graham, Acting Deputy Director, Agency for Healthcare Research and Quality, Rockville, Md.; Lee Greenfield, Policy Adviser, Hennepin County Health and Community Integration, Minneapolis, Minn.; Jessie Gruman, President and Executive Director, Center for the Advancement of Health, Washington, D.C.; Gordon Guyatt, Professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Ontario, Canada; Mark Helfand, Associate Professor of Medicine, Division of Medical Informatics and Outcomes Research, Oregon Health and Science University, Portland; David Henry, Professor, School of Medical Practice and Population Health, University of Newcastle, Callaghan, Australia; Chris Henshall, Director, Science and Engineering Base Group, Office

of Science and Technology, Department of Trade and Industry, London, U.K.; Roger Herdman, Director, National Cancer Policy Board, Institute of Medicine, Washington, D.C.; Clarence J. Hindman, Chief Medical Officer and Interim Medicaid Director, Arizona Health Care Cost Containment System, Phoenix; Hans V. Hogerzeil, Acting Director, Department of Essential Drugs and Medicine, World Health Organization, Geneva, Switzerland; Walter Holland, Professor, Health and Social Care, London School of Economics and Political Science; Andrew Holtz, President, Association of Healthcare Journalists, Portland, Ore.; Betsy Humphreys, Associate Director, Library Operations, National Library of Medicine, National Institutes of Health, Bethesda, Md.; Tom Huntley, Lead Democrat, Health and Human Services Finance Committee, Minnesota House of Representatives; Clarion E. Johnson, Global Medical Director, Exxon Mobil Corporation, Fairfax, Va.; Dennis Kendel, Registrar, College of Physicians and Surgeons of Saskatchewan, Canada; Rudolf E. Klein, London School of Hygiene and Tropical Medicine and London School of Economics; Kurt Knickrehm, Director, Arkansas Department of Human Services, Little Rock; Allan Korn, Senior Vice President and Chief Medical Officer, Blue Cross Blue Shield Association, Chicago; Joseph Lau, Director, Division of Clinical Care Research, Tufts–New England Medical Center, Boston; Anthony F. Lehman, Professor and Chair, Department of Psychiatry, University of Maryland School of Medicine, Baltimore; Jeffrey Lerner, President and Chief Executive Officer, ECRI, Plymouth Meeting, Pa.; Ian Leverton, Vice President, Clinical Integration, Sutter Health, Sacramento, Calif.; Sharon Levine, Associate Executive Director, The Permanente Medical Group, Oakland, Calif.; Steven Lewis, President, Access Consulting, Saskatchewan, Canada; Jonathan Lomas, Executive Director, Canadian Health Services Research Foundation, Ontario, Canada; Malcolm Maclure, Professor, School of Health Information Science, University of Victoria, British Columbia, Canada; Nicholas Mays, Professor of Health Policy, London School of Hygiene and Tropical Medicine; Anne McFarlane, Assistant Deputy Minister, Planning, Policy and Legislation Branch, Ministry of Health Planning, British Columbia, Canada; Vin McLoughlin, Assistant Secretary, Health Priorities Branch, Department of Health and Ageing, Canberra, Australia; Stewart McMillan, Chair, Department of Family Medicine, Regina-Qu'Appelle Health Authority, Saskatchewan, Canada; B. Joyce McRoberts, Deputy Director, Idaho Department of Health and Welfare, Boise; Velvet G. Miller, President and Chief Executive Officer, My Parent's Concierge, Trenton, N.J.; Edward Mills, Director of Research, Canadian College of Neuropathic Medicine, Toronto; Kenneth C. Montague, Secretary, Maryland Department of Juvenile Justice, Baltimore; Victor Montori, Professor, McMaster University, Ontario, Canada; Cynthia D. Mulrow, Professor, Department of Medicine, University of Texas Health Science Center, San Antonio; Robert Nakagawa, Director of Pharmacy, Fraser Health Authority, British Columbia, Canada; Maryann Napoli, Associate Director, Center for Medical Consumers, New York City; Andrew Oxman, Director, Department of Health Services Research, Norwegian Directorate for Health and Social Welfare, Oslo; Tonse Raju, Medical Officer, Pregnancy Perinatology Branch, National Institute of Child Health and Human Development, National Institutes of Health, Rockville, Md.; Daryl Rock, Director, Strategic Programs and Joint Initiatives, Social Sciences and Humanities Research Council, Ontario, Canada; Peggy Rosenzweig, Wauwatosa, Wisc.; Jill M. Sanders, President,

Canadian Coordinating Office for Health Technology Assessment, Ontario; Paul Shekelle, Professor, Department of Medicine, University of California at Los Angeles; Yvonne M. Shevchuk, Division Head and Professor, College of Pharmacy and Nutrition, University of Saskatchewan, Canada; Jean Slutsky, Acting Director, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, Rockville, Md.; Marlene Smadu, Associate Dean of Nursing, College of Nursing, University of Saskatchewan, Canada; Alison Spaul, Director, Chief Scientist Office, The Health Department, Edinburgh, U.K.; Ronald J. Stamp, Professor, University of Portsmouth, Hampshire, U.K.; Serge Taillon, Director, Client Relations and Communications, Canadian Institute for Health Information, Ontario, Canada; Octavi Quintana Trias, Directorate-General for Research, Life Sciences, Research Actions for Health, Brussels, Belgium; Peter Tugwell, Director, Center for Global Health, University of Ottawa, Canada; Sean Tunis, Chief Medical Officer, Centers for Medicare & Medicaid Services, Baltimore, Md.; Jed Weissberg, Associate Executive Director, Quality and Performance Improvement, The Permanente Federation, Oakland, Calif.; Terri Wetle, Professor, Department of Biology and Medicine, Brown University, Providence, R.I.; Michael Whitcomb, Senior Vice President, Division of Medical Education, Association of American Medical Colleges, Washington, D.C.; Thomas R. Young, Medicaid Director, Idaho Department of Health and Welfare, Boise.

EXECUTIVE SUMMARY

The systematic review is now widely regarded as the least biased and most rational way to summarize the research evidence that evaluates health care interventions meant to prevent and treat illness. A systematic review can help distinguish therapies or interventions that work from those that are useless, harmful, or wasteful. It can reliably estimate how well different options work, and it can identify gaps in knowledge requiring further research.

Efforts to synthesize research evidence date back centuries, but the science of the systematic review has been greatly refined in the past few decades, particularly in the social and health sciences. As Richard Light and David Pillemer explain in their groundbreaking work *Summing Up: The Science of Reviewing Research*, “Without a clear picture of where things stand now, simply adding one new study to the existing morass is unlikely to be very useful. . . . For science to be cumulative, an intermediate step between past and future research is necessary: synthesis of existing evidence.”

The basic steps of a systematic review include formulating a question; finding relevant studies; selecting and assessing the studies; summarizing and synthesizing study results; interpreting the review results; and maintaining and updating the review.

Systematic reviews are being produced by many organizations in the United States and around the world, and they are being used across much of the health care landscape. Growing numbers of nurses, pharmacists, doctors and other health professionals, patients, insurers, policymakers, advocates, health care executives, and journalists are looking to systematic reviews of the evidence to inform their thinking, decision making, and practice. But systematic reviews have significant limitations as well as benefits. The studies being reviewed are often incomplete, deficient, or skewed toward the most profitable treatments. Sometimes systematic reviews themselves are poorly conducted, as for example when the search for relevant studies has not been as comprehensive as possible. Often systematic reviews will conclude that there is not enough strong evidence to support or refute a technology that some clinicians and patients consider promising—yet decisions about its use must still be made and defended.

Notwithstanding their pitfalls, systematic reviews promise to improve the quality of many health care decisions.

A NOTE ABOUT NOMENCLATURE

Research synthesis in health care is an evolving science. It is an old practice whose methods have been refined in recent decades within the social and health sciences.

The most widely used term for the process is *systematic review*: First a clear question is formulated—like “Do antidepressants help smoking cessation?” Then specific methods are used to identify, select, and appraise relevant research. Finally, the results are analyzed and summarized.

Another commonly used term is *meta-analysis*, the application of the statistical method for pooling and synthesizing the quantitative results of different studies. A meta-analysis may stand alone or be included within a systematic review.

Currently the results of systematic reviews and meta-analyses are presented in different formats and under different titles, depending on the organizations involved. Internationally the Cochrane Collaboration produces *Cochrane Reviews*, which are individual systematic reviews of the evidence about the effects of particular health care interventions. In the United States and Canada, the Evidence-Based Practice Centers produce *Evidence Reports* and *Technology Assessments*, which generally focus on broader questions than individual Cochrane Reviews. These publications may include a summary of earlier systematic reviews or analyze cost-effectiveness. The International Network of Agencies for Health Technology Assessment publishes *Systematic Reviews*, *Economic Evaluations*, and *Synthesis Reports* sourced to agencies in 20 countries. The World Health Organization’s Health Evidence Network produces *Synthesis Reports* in response to broad questions from health care decision makers across Europe.

The principles underpinning all these terms are very simple: Producing and using systematic reviews of the best evidence can help us all make more informed decisions about health care.

For simplicity’s sake, this report will usually use the term *systematic review*.

INTRODUCTION TO SYSTEMATIC REVIEWS

If we want to base our health care decisions on evidence, it makes sense to use the best evidence available. By and large the best evidence for many decisions will come from a systematic review of all the evidence. A specific and reproducible method is used to identify, select, and appraise all the studies that meet minimum quality standards and are relevant to a particular question. The results of the studies are then analyzed and summarized. It sounds like a simple idea because it is—simple but powerful.

Systematically reviewing scientific evidence about health care is a way of finding out what we know and don't know about what works and what doesn't work. A good systematic review can generally give us the most reliable estimate of the effectiveness of a specific intervention, and it can identify gaps in our knowledge that require further research. It can also give us a sense of the strength of the available evidence and the quality of the studies, thereby indicating how much confidence clinicians, patients, policymakers, and journalists should have in the results. (See Box 1, page 4.)

A systematic review can also at times help us find out how well a treatment works for different kinds of patients and inform us about its potential to cause harm. Some systematic reviews can shed light on the pros and cons of different ways of organizing or delivering care. A systematic review that includes considerations of cost can help shape our judgments about whether particular technologies or policies provide good value for money.

Like all science, the science of research synthesis is evolving and uncertain. For example, the application of the statistical method for pooling and synthesizing the quantitative results of different studies—called meta-analysis—is steadily improving, though considerable challenges remain. Another problem is that the potential side effects of a treatment will not be adequately reflected in a systematic review if they were inadequately investigated in the original studies.

A good systematic review frequently highlights weaknesses in the evidence for a therapy's benefits and argues for further research, creating dilemmas for patients, policymakers, and others who must make decisions today. Even when a systematic review shows strong, unambiguous evidence to support one course of action, the politics of health care may make that review's findings less influential than a well-publicized series of anecdotes or the advocacy of a different course by a politically savvy group, corporation, or trade association. Yet even if it is overlooked, a systematic review of the evidence can allow more transparency about the role of other factors in decision making. No matter how well they are done, systematic reviews are not a panacea for all problems, but they sometimes offer a valuable aid to decision making.

As a general rule, weak study designs or poor-quality studies can produce misleading findings that are less reliable and less accurate, often tending to overstate or exaggerate a therapy's benefits. (See Box 2, page 11.) The critical scrutiny central to a systematic review can help point out those weaknesses and identify more accurate, more reliable evidence.

Traditional reviews of the health care literature—sometimes called narrative reviews—are by definition nonsystematic: They lack an explicit and reproducible method for identifying, selecting, and assessing the studies being reviewed. While traditional reviews may be appropriate and valuable for some

BOX 1. LEVELS OF EVIDENCE

Different types of study designs produce different levels of evidence. Some study designs produce weaker forms of evidence; other designs produce stronger, less biased, and thus more reliable evidence. Over the last two decades, medical scientists have developed a “hierarchy” of evidence. Several such hierarchies have been produced. The international GRADE Working Group is developing a common approach to grading evidence. A simple hierarchy, from the strongest level to the weakest, is shown here, based on one recently produced by the Centre for Evidence-Based Medicine at Oxford University:

Systematic Reviews of Randomized Controlled Trials
Randomized Controlled Trials
Observational Studies
Case Series
Expert Opinion

This hierarchy is complex and evolving, not least because many research questions cannot be studied with randomized controlled trials, in which study participants exposed to an intervention are compared to participants in a control group who are not exposed to that intervention. For example, to study the long-term effects of smoking, people cannot be randomly assigned to a smoking group and a nonsmoking group. Instead researchers use observational studies, where the health of smokers is compared to that of a similar group of nonsmokers. Much research about the organization or delivery of health services is also not amenable to randomized controlled trials. The important point is that the study question will influence the choice of study design, so the hierarchy shown here may vary, depending on the study question.

The other major caveat is that research at any level in the hierarchy, including systematic reviews, can be of high or low quality. For instance, randomized controlled trials where both patients and investigators are “blinded” to whether certain patients are receiving the treatment or not are generally considered better quality than trials that are not blinded. (Sometimes blinding is not possible or appropriate.)

Explanations of different study designs, different levels of evidence, and the methods for appraising the quality of a study are beyond the scope of this report. Readers interested in learning more can contact any of the organizations listed in Box 3, page 18.

subjects and topics, systematic reviews are widely seen as the best way to answer questions about how well health care interventions work. For example, a recent systematic review summarized the results of 20 studies of the benefits of smoking cessation for people with heart disease. It found that quitting smoking reduced the risk of fatal and nonfatal heart attacks among these people by more than 30 percent compared to those who continued smoking.

While many medical technologies, tests, and procedures have been proven effective and safe, many others are used in the absence of good-quality scientific evidence, despite being recommended by medical “experts,” who are often the authors of traditional nonsystematic reviews. When rigorous scientific evaluation does belatedly take place, widely used interventions are sometimes shown to be ineffective or potentially harmful, as was the case with the long-term use of combination hormone therapy.

Organizations in the United States and elsewhere are now producing systematic reviews of the evidence about medical therapies, prescription medicines, diagnostic and screening tests, surgical procedures, disease incidence, and prevalence strategies. Researchers are also producing systematic reviews of the evidence about different ways to organize and deliver care and other policy and management-oriented interventions. (See Box 3, page 18.) While much of this report is concerned with systematic reviews of evidence about health care interventions—drugs, therapies, technologies—the principles of systematic reviewing remain the same whatever subject matter is being reviewed.

The reviews come in different formats under different names—systematic reviews, evidence reports, technology assessments, overviews, meta-analyses, Cochrane Reviews, etc.—but they all share a commitment to the rigorous review of the evidence. Many of these reviews are now available on the Web, some with user-friendly summaries attached.

There is anecdotal evidence that different players in the health care sector are using systematic reviews in their day-to-day decision making, whether the players are public or private sector policymakers, insurers, hospitals, clinical guideline writers, health care professionals, consumers, or journalists. Some national systems and numerous health plans now routinely require systematic reviews before making coverage decisions about expensive new therapies; so do Medicare and some state Medicaid programs in the United States. But systematic reviews are meant to assist the process of decision making, not replace it. And for many complex health care decisions, good systematic reviews will not be available or applicable.

The history of synthesizing research is inextricably bound up in the history of evidence-based medicine—the global movement to use the best evidence about what does and does not work in health care. A perfect introduction to this dual history comes from the work of Scottish naval surgeon James Lind, who is credited with having produced one of the early records of a scientific trial and having written one of the first systematic reviews of evidence.

On board the *Salisbury* on May 20, 1747, Lind took 12 patients with scurvy, whose cases “were as similar as I could have them.” He divided them into six groups of two and administered different treatments to each pair of sufferers. The six treatments were cider, elixir vitriol, vinegar, seawater, a combination of oranges and lemons, and a mixture of garlic, mustard seed, and balsam of Peru. Six days later, Lind’s findings were clear: “The result of all my experiments was that oranges and lemons were the most effectual remedies for this distemper at sea.”

Six years later, the results were published under the following title: *A Treatise of the Scurvy in Three Parts. Containing an inquiry into the Nature, Causes and Cure of that Disease, together with a Critical and Chronological View of what has been published on the subject*. Acknowledging the need to review the existing literature on scurvy systematically and to discard the weaker forms of evidence, Lind wrote: “As it is no easy matter to root out prejudices . . . it became requisite to exhibit a full and impartial view of what had hitherto been published on the scurvy . . . by which the sources of these mistakes may be detected. Indeed, before the subject could be set in a clear and proper light, it was necessary to remove a great deal of rubbish.” Gathering the published research, getting rid of the rubbish, and summarizing the best of what remains is essentially the science of systematic review.

Through the early decades of the 20th century, scientists working in diverse areas from environmental air quality to physics and agriculture employed rudimentary techniques of research synthesis. In 1904 Karl Pearson, director of the Biometric Laboratory at University College London, published a landmark review of the evidence about the effects of vaccines against typhoid. Pearson summarized and synthesized the results of 11 studies, thereby anticipating the development of meta-analysis, the statistical method used to pool the results of different studies. Three years later a scientist in the United States reviewed 44 studies of typhoid fever in the District of Columbia. Identifying the most reliable studies, Joseph Goldberger then abstracted and pooled data from 26 of the 44 studies.

During the 1960s and 1970s, social scientists further developed and refined the methods for systematically reviewing large bodies of evidence, summarizing results from hundreds of different studies on the same topic. Their subjects varied widely and included the effects of class sizes on academic achievement, the effectiveness of psychotherapy, and the validity of employment tests. The twin goals of the process, then as now, were to minimize bias in the final review by discarding or criticizing weak studies and to maximize confidence about how well an intervention works by summarizing the findings from the strongest and best studies.

As the methods for systematic reviews were developing, so too were the scientific methods for conducting the studies under review. The growth of controlled trials led to the emergence of randomized controlled trials (RCTs) in the 1940s. The RCT, like the systematic review, is based on a

simple but powerful idea. Similar patients are randomly assigned to groups; one group receives the treatment or intervention; the “control” group does not. If more than one treatment is being tested, more groups are involved. Random assignment is intended to create groups of participants as similar to each other as possible at the start of the experiment, so that any differences at the end can be attributed to the intervention being studied. Researchers who choose which group to assign their patients to may bias the experiment by consciously or unconsciously assigning particular kinds of patients to particular groups.

In 1971 Archie Cochrane, a British epidemiologist, published his influential book *Effectiveness and Efficiency: Random Reflections on Health Services*. He persuasively advocated the scientific evaluation of commonly used medical therapies through randomized controlled trials, essentially foreshadowing the rise of evidence-based medicine. By 1979 Cochrane was suggesting that the results of randomized controlled trials of the same intervention be systematically summarized. This time he was anticipating the creation of a global collaboration that would begin to produce such summaries just over a decade later.

In 1984 Richard Light and David Pillemer published *Summing Up: The Science of Reviewing Research*, a book widely regarded as a pioneering work in the recent history of research synthesis. “Our broad goal,” they wrote, “is to help readers organize existing evidence in a systematic way, whether a review is motivated by a scientific problem or the need for a policy decision.” They argued that the new methods of systematically reviewing evidence could be applied to many fields, including health, education, and psychology.

Three years later the science of research synthesis leapt forward when Cynthia Mulrow, a University of Texas researcher, delivered her damning assessment of the quality of 50 reviews published in the world’s leading medical journals during 1985 and 1986. Using eight criteria, she found that only one of the 50 reviews “had clearly specified methods of identifying, selecting, and validating included information.” She concluded: “Current medical reviews do not routinely use scientific methods to identify, assess, and synthesize information.” On the contrary, these reviews were often “subjective, scientifically unsound, and inefficient.”

The exponential growth in medical literature during the 1980s made the need for systematic reviews of evidence more and more important. At the same time, rapidly growing health care expenditures made it increasingly critical to assess whether expensive new technologies offered value for money. Systematically reviewing evidence of the benefits and harms of new therapies would become a key part of the emerging science of cost-effectiveness, or cost-benefit analysis.

Mulrow’s proposals for a more systematic approach to reviewing and summarizing medical evidence were developed further by Andrew Oxman, Gordon Guyatt, and other colleagues at McMaster University in Hamilton, Ontario, Canada. In 1993 Oxman and Guyatt published their own iconoclastic critique of the poor quality of review articles, based on an assessment of 36 published reviews. They surveyed the reviews’ authors about their levels of expertise, the time they had spent on their reviews, and the strength of their prior opinions. Then they convened a panel of researchers to

judge the methodological rigor of each of the 36 reviews, using criteria like whether the reviewer had specified his or her methodology, and whether the review's conclusions were supported by the reported data. Oxman and Guyatt concluded: "Our data suggest that experts, on average, write reviews of inferior quality; that the greater the expertise the more likely the quality is to be poor; and that the poor quality may be related to the strength of the prior opinions and the amount of time they spend preparing a review article."

A key element of the history of systematic reviewing is the development of methods for critically appraising scientific studies. One milestone was a 1979 report of the Canadian Task Force on the Periodic Health Examination. It applied a new system for rating the strength and quality of studies. The task force report analyzed studies of interventions for many common medical conditions and rated the evidence according to a now familiar hierarchy, with randomized controlled trials at the top, observational studies in the middle, and experts' opinions at the bottom. (See Box 1, page 4.)

Through the latter part of the 1980s and into the 1990s, the "outcomes movement" also stimulated the demand for systematic reviews. Motivated in part by the large variations in the use of medical interventions for the same conditions in different towns, cities, and regions, a federal agency began funding academic teams across the United States to evaluate the effects of common treatments and to develop guidelines for the management of common conditions, including back pain, benign prostate disease, cataracts, and knee problems. These Patient Outcome Research Teams (PORTs) would systematically review the medical evidence supporting widely used therapies.

These teams often found a lack of high-quality scientific studies. Reviewing the evidence regarding prostate surgery for benign prostatic hyperplasia (an enlarged prostate gland in older men), for example, a PORT led by Dartmouth Medical School researcher Jack Wennberg could find almost no randomized controlled trials of the procedure. The PORT examining total knee replacement searched the medical literature and found only a few small studies of extremely poor quality. The PORT concerned with treatments for chronic lower-back pain concluded that all the relevant scientific studies provided insufficient evidence to justify the widespread use of spinal fusion surgery. A subsequent guideline that used this finding helped spark a backlash from surgeons and device-manufacturers that, for a time, threatened congressional funding of the Agency for Health Care Policy and Research, now called the Agency for Healthcare Research and Quality. As Iain Chalmers, editor of the *James Lind Library*, and his colleagues report in a brief history of this science, "Research synthesis sometimes yields unwelcome results that challenge strongly held opinions and other vested interests."

In the United Kingdom through the 1980s, Chalmers led a project to systematically review the scientific evidence for all interventions and procedures related to pregnancy and childbirth. The resulting publication summarized the evidence for what worked and what didn't, listing the many effective procedures alongside dozens of forms of care that were useless, harmful, or both. The project's methods became a template for the methods adopted soon after by an international group called the Cochrane Collaboration, organized by Chalmers and others in 1993 and named for Archie Cochrane, who had died in 1988.

The mission of the Cochrane Collaboration is to produce systematic reviews of the evidence about all health care interventions. Starting with just a few Cochrane centers in a handful of countries and largely supported by public funds, the collaboration has quickly built on the hard work of thousands of researchers and consumers around the world. Its electronic database, the Cochrane Library, now boasts almost 2,000 systematic reviews that are regularly updated, as well as a register of more than 350,000 trials and other studies. Abstracts of those reviews are available free on the Web, with this accompanying explanatory note: “Cochrane Reviews make the results of research assessing the effects of health care more easily available to those who want to make better decisions. . . . Well-informed decisions also require information, and judgments about needs, resources and values. . . . Care and compassion are vital, and understanding the nature and basis of disease and the way that interventions work remains important.” Governments in a number of countries, including Australia, Ireland, Norway, Spain, and the United Kingdom, and the Canadian province of Saskatchewan have decided to subsidize free access to the full text of Cochrane Reviews for their citizens.

Many other health care organizations began producing systematic reviews during the 1990s in the United States and elsewhere around the world, and many groups in both the private and public sectors were promoting the need for such evidence. In 1995 Australia’s federal government mandated that pharmaceutical companies submit a systematic review of the relevant biomedical literature when applying for listing of their products in the national formulary. A number of national research funding agencies in other countries, including the United Kingdom and the Netherlands, began requiring systematic reviews of existing research before considering applications for funding for further study. The national research ethics committee in Denmark went a step further, requiring applicants for new studies of primary data to “show by reference to syntheses of existing evidence that proposed new studies were necessary and that they had been designed to take account of the lessons from previous research.”

In the late 1990s, the Agency for Healthcare Research and Quality in the U.S. Department of Health and Human Services designated research groups in the United States and Canada as Evidence-Based Practice Centers. These centers, which now number 13, conduct “systematic, comprehensive analyses and syntheses of the scientific literature to develop evidence reports and technology assessments on clinical topics that are common, expensive, and present challenges to decision makers.” In 1999 the U.S. agency responsible for Medicare announced that it would require such systematic reviews of the evidence before making any major national coverage decisions, though this policy does not affect the many coverage decisions about therapies made at the regional or state level.

WHY DO SYSTEMATIC REVIEWS?

There are many good reasons to review health care evidence systematically. Systematic reviewing is currently the best, least biased, and most rational way to organize, cull, evaluate, and integrate the research evidence from among the expanding medical and health care literature. The results of a systematic review can help distinguish therapies and interventions that work from those that are useless or harmful and can replace guesswork with more reliable estimates of how well things function. A systematic review can also identify what is known and what is unknown, giving guidance for further research. As Light and Pillemer explain in *Summing Up*: “Without a clear picture of where things stand now, simply adding one new study to the existing morass is unlikely to be very useful. . . . For science to be cumulative, an intermediate step between past and future research is necessary: synthesis of existing evidence.”

Cynthia Mulrow, University of Texas professor and deputy editor of the *Annals of Internal Medicine*, argues that reviewing in this way is a search for the whole truth, rather than just one part of it, and is thus a “fundamentally scientific activity.” In 1998 Mulrow cowrote a guide to systematic reviewing, aimed primarily at the practicing clinician and published by the American College of Physicians. Picking up on the widespread enthusiasm for the human genome project, she wrote: “Exciting new information pouring out of the molecular biology revolution has the potential to transform medicine. But even this enormously powerful information will be of little use to physicians and their patients unless (1) the diagnostic and therapeutic interventions that flow from it are stringently tested in clinical trials and (2) the results of those trials are synthesized and made accessible to practitioners. . . . Systematic reviews are thus a vital link in the great chain of evidence that stretches from the laboratory bench to the bedside.” In the opening chapter of *Systematic Reviews*, Mulrow briefly points out the main reasons for systematically reviewing health care evidence. Reviewing and synthesizing is a way to sort through the growing mountain of data. There are now millions of studies in the health care literature, and systematic reviewing helps separate “the insignificant, unsound, or redundant deadwood from the salient and critical studies that are worthy of reflection.”

It is important to identify studies with weak designs or poor quality, because their results are frequently biased and misleading, often overestimating the benefits of the treatment being studied. (See Box 2, page 11.) Once these studies have been identified, the more reliable evidence can be summarized to make it useful for health professionals, policymakers, consumers, and others.

Some health services researchers emphasize that the value of systematic reviewing is not limited to the results of quantitative clinical studies. Nicholas Mays of the London School of Hygiene and Tropical Medicine, and colleagues, for instance, recently outlined methods for synthesizing research evidence about the organization and delivery of health services, which comes largely from nonexperimental or qualitative studies. Regardless of the nature of the evidence, says Mays, health care managers, policymakers, and professionals will value “overviews which succinctly bring together a potentially large and often scattered body of research in a manageable way for quick absorption.”

Another reason for systematically reviewing many different studies of the same question is that the review can produce evidence that is more “generalizable,” that is, potentially applicable to a wide

range of patients or settings. The results of a single study often apply only to a certain kind of patient or a particular policy setting, like a hospital, for example. A systematic review of many studies can provide information relevant to a broad range of patients at different treatment doses in different treatment settings. Clearly, applying the results of any study or systematic review to the particular problems of individual patients in a clinical setting will remain a key challenge, no matter how strong the evidence is.

Yet another reason to synthesize the results of different studies of a given intervention is to learn how consistent the studies' findings are. Light and Pillemer have written that "disagreements among findings are valuable . . . [and that] conflicts can teach us a lot," because we can see in what settings a particular educational program might work best or what dose of a drug might work best.

DANGEROUS DISCREPANCIES BETWEEN EXPERTS AND EVIDENCE

In July 1992, the *Journal of the American Medical Association (JAMA)* published the results of an extraordinary study, now cited regularly as one of the strongest justifications for producing and using systematic reviews of evidence, rather than relying on the recommendations of experts and their nonsystematic reviews of the literature. A team of researchers from several institutions in Boston set

BOX 2. WEAKER STUDIES TEND TO EXAGGERATE AND MISLEAD

Several scientific articles have published empirical evidence of bias in weaker studies. The articles have shown that studies with weaker designs or poor quality frequently tend to overstate benefits—thus highlighting the need for stronger study designs, better-quality studies, and systematic reviews of those studies:

- Nonrandomized trials of anticoagulant medicines exaggerated their benefits compared to randomized trials.
- Nonblinded studies of low-molecular-weight heparin exaggerated its benefits compared to blinded studies.
- Nonblinded studies of pregnancy and childbirth therapies exaggerated their benefits compared to double-blinded studies.
- Poorly randomized trials of pregnancy and childbirth therapies exaggerated their benefits compared to well-randomized trials.
- Low-quality studies of interventions used for circulatory and digestive diseases, mental health, and pregnancy and childbirth exaggerated the interventions' benefits compared to high-quality studies.
- A systematic review of all the relevant evidence suggested that on average, weaker studies tend to overstate benefits, but this is not always the case.

out to compare the evidence from randomized controlled trials of treatments for heart attacks with the recommendations of clinical experts published in review articles, textbooks, and manuals.

Using a statistical method called “cumulative meta-analysis,” Antman and colleagues looked at the latest accumulated evidence for every year between 1960 and 1990 about the effectiveness of commonly used treatments to reduce the risk of heart attack—including thrombolytic therapy, prophylactic lidocaine, class I antiarrhythmics, and several others. They then compared the latest results to what “experts,” “opinion leaders,” or “thought leaders” were recommending in books and review articles in that year to see whether they were recommending routine use, specific use for certain patients, or no use.

The *JAMA* study found major discrepancies between the accumulating evidence and the experts’ recommendations. In most instances where studies showed treatments to be effective, experts’ recommendations lagged several years behind the evidence. The most notable example was thrombolytic drugs (used to break up blood clots): They were not recommended by more than half the experts until 13 years after the cumulative evidence showed them to be effective. Moreover it took six years after the first published meta-analysis showed these drugs to be effective before a majority of experts recommended their routine or specific use.

Regarding the use of lidocaine to prevent ventricular fibrillation—that is, to maintain the beating of the heart—the *JAMA* study found that most experts over a 25-year period recommended use of the drug, although controlled studies provided no evidence that it reduced deaths. In another example, a small number of experts were still recommending long-term use of antiarrhythmic drugs (to correct irregularities in the heartbeat) at a time when the drugs had been shown to increase the death rate.

These two examples reveal the difference between a lack of evidence of effectiveness and evidence of a lack of effectiveness. With lidocaine, evidence of effectiveness was lacking—which is not the same as evidence of ineffectiveness. With long-term use of antiarrhythmic drugs, evidence of ineffectiveness was available; indeed strong evidence showed that the drugs harmed or killed many people.

Thomas Moore, a George Washington University fellow and investigative journalist, has shown that widespread use of antiarrhythmic drugs, including encainide and flecainide, ultimately caused the deaths of tens of thousands of Americans. While a systematic review of the evidence suggested serious uncertainties about the benefits of these medicines as early as 1983, this review appears to have been overlooked. The drugs were widely used until the 1990s. For Iain Chalmers, the deaths caused by these drugs are a compelling reason to take systematic reviews of evidence seriously and to place new evidence from single studies in the context of what is already known from the systematic reviews of previous evidence: “The warning signs were there in one of the first systematic reviews of controlled trials in health care, yet more than 50 trials of these drugs were conducted... before official warnings about their lethal impact were assessed. Had the new data generated by each one of these trials been presented within the context of systematic reviews of the results of all previous trials, the lethal potential of this class of drugs would have become clear earlier, and an iatrogenic [a condition or disease caused by treatment] disaster would have been contained, if not avoided.”

One of the earliest individual trials of antiarrhythmic drugs found that they increased the death rate, but the study remained unpublished for two decades. This long delay is an example of the wide problem of “publication bias”: Studies that find negative or unfavorable results are far less likely to be published than studies with positive findings. As we will see in the next section, a good systematic review of evidence will search for unpublished as well as published studies.

Antman and colleagues summarized the findings of their landmark study, writing that some experts “have not yet mentioned effective therapies, while others continue to recommend those that are ineffective or possibly harmful.” They concluded by suggesting that meta-analyses that pool the findings of high-quality trials could help opinion leaders and regulatory bodies synthesize the burgeoning literature and help inform choices for appropriate therapies.

A final reason to do systematic reviews is offered by those who purchase health care. As president of the National Business Group on Health, in Washington, D.C., Helen Darling represents 175 large corporations that offer health insurance coverage to approximately 40 million employees, dependents, and retirees in the United States. “There are constant financial incentives in this system to do more of everything and do it at the most lavish level,” Darling notes. “Our incentive is first and foremost to make certain any new technology is worth paying for. The only way to answer that question is to start with a systematic review of the evidence.”

HOW TO DO A SYSTEMATIC REVIEW

People who conduct systematic reviews are explicit about how they have tried to search for all the relevant studies, how those studies are assessed, and how their results are summarized. With an explicit, reproducible method, it becomes much harder for reviewers to gather or summarize only studies that support their biases.

Different manuals enumerate different steps for a review. There are, perhaps, six key steps, based on those outlined in *Systematic Reviews in Health Care: A Practical Guide*, written jointly by researchers in Australia and the United States.

1. Formulating a question.
2. Finding relevant studies.
3. Selecting and assessing those studies.
4. Summarizing and synthesizing relevant study results.
5. Interpreting the review findings.
6. Updating the review.

FORMULATING A QUESTION

The first step, formulating a clear, apt question for the review to answer, is often difficult. Reviews can address many different types of questions about various aspects of the health care system, not just “How well does this therapy work?” or “How does therapy A compare to therapy B?” Researchers can also review the evidence about diagnostic tests and screening programs; risk factors for certain illnesses; nonmedical approaches; disease prevalence; and broad policy or management issues that affect the organization or delivery of health services.

Different questions will spark a search for different kinds of evidence from different types of primary or original studies. A systematic review about the effectiveness of prescription drugs will search for evidence from top-quality randomized controlled trials. A systematic review about the effects of smoking will find no trials that assigned some people to smoke and will therefore seek observational studies, which compare the health of similar groups of smokers and nonsmokers.

Another key consideration is the breadth of the question. As the *Practical Guide* explains, asking “What is the effect of cancer screening on the general population?” is too broad. Conversely, “What is the mortality reduction in colorectal cancer from yearly fecal occult blood screening in 40-to-45-year-old females?” is too narrow. A better question for a review might be, “What is the mortality reduction in colorectal cancer from fecal occult blood screening in adults?”

The formulation of a question can be an iterative approach, which may develop during the initial literature searches. If researchers are interested in addressing questions relevant to policymakers, managers, health professionals, or consumers, some sort of consultation process may be useful while the question is being formulated.

FINDING RELEVANT STUDIES

Once a question has been formulated, the next step is finding studies. As the reviewers' handbook used by the Cochrane Collaboration points out, "Conducting a comprehensive, objective, and reproducible search for studies can be the most time-consuming and challenging task in preparing a systematic review. Yet it is also one of the most important. Identifying all relevant studies, and documenting the search for studies with sufficient detail so that it can be reproduced is, after all, largely what distinguishes a systematic review from a traditional narrative review." The details of search methods are explained in many other publications. In brief they involve entering explicit search terms into electronic databases such as CENTRAL, EMBASE, and MEDLINE; searching older journals and "gray" or "fugitive" literature, including conference proceedings, by hand; checking the reference lists of retrieved studies; and writing to colleagues and companies with an interest in the topic to seek relevant unpublished studies.

If reviewers are searching for studies across a range of different academic perspectives, or across a range of methodologies that includes both qualitative and quantitative designs, according to Nicholas Mays, reviewers may not be able to settle on a fixed search strategy. They may need to be flexible and to modify the strategy as the search unfolds. For Mays, the important point is to document the twists and turns in the search, so that others can repeat it: "Comprehensiveness, explicitness about methods and reproducibility should also be the goal of researchers synthesizing evidence from service delivery and organisational research, in order to limit bias and make conclusions more reliable."

Seeking unpublished studies is particularly important. Because studies that find no effect for a treatment or intervention are much less likely to be published, relying only on published trials can lead to an overestimate of the benefits of health care interventions. In 1986 the Australian researcher John Simes demonstrated the problem of "publication bias" with two examples involving cancer therapies.

In the first example, Simes compared two reviews of the evidence about chemotherapy drugs for ovarian cancer. One review summarized only published trials; the other summarized unpublished studies as well by including studies from a register of cancer trials. The more limited review found combination therapy more effective than single therapy at extending survival. The more extensive search, which included trials from the register, found no benefit for combination therapy.

In the second example, involving multiple myeloma (disease of the bone marrow), the review limited to published trials found combination therapy more effective than single therapy. The more extensive review that included unpublished trials also found combination therapy more effective than single therapy, but the estimated magnitude of the extra benefit was smaller.

In both examples, the reviews of only the published literature produced inflated estimates of the therapies' benefits when compared to the more comprehensive reviews. Simes's findings have been echoed in other studies of publication bias for almost two decades.

SELECTING AND ASSESSING THOSE STUDIES

An explicit and standardized method for selecting studies from among all of those identified and then assessing the selections is a key part of a systematic review. Such a method serves the dual purpose of choosing the highest-quality studies and demonstrating to readers that the selection and assessment have been as free from bias as possible.

According to Glasziou and colleagues, there are two parts to the selection process. The initial screen excludes studies not relevant to the review question. The second screen determines which of the relevant studies are strong enough to be included in the systematic review.

There are many different scales for assessing the strength and quality of a study that may be included in a review. Like the whole science of research synthesis, these scales are evolving. However, there is a consensus that at least two reviewers should independently select and assess studies for inclusion or exclusion, according to explicit criteria, and that discrepancies should be resolved by discussion.

The Agency for Healthcare Research and Quality recently examined over 100 different scales for assessing the strength and quality of a study or review. The resulting report summarized the findings with a checklist of commonly used criteria and suggested a number of scales that included many of the most important criteria. For example, one scale deemed to be valuable was that used by Deborah Barnes and Lisa Bero while judging the quality of review articles on the health effects of passive smoking.

The way that studies are assessed can determine which studies are included in any quantitative summary of results and can greatly affect the findings of a systematic review. In 1999 Peter Juni and colleagues took an existing meta-analysis of 17 trials of a drug used to reduce the risk of deep vein thrombosis. They then conducted more meta-analyses of the same 17 trials—each time using a different scale for assessing their quality. They found that using different scales led to different findings in the meta-analyses.

Similarly, part of the controversy surrounding the systematic review of mammography screening for breast cancer, published in the *Lancet* in 2001, arose because different researchers have different views on the quality and strength of the original studies that were reviewed. Disagreements over approaches are likely to persist, despite international efforts to reduce them.

SUMMARIZING AND SYNTHESIZING RELEVANT STUDY RESULTS

One of the chief goals of a systematic review of the evidence is to summarize the findings of the best studies available. A concise written summary of each of these relevant studies is usually provided, often as a table of summaries, which is called a summary table or evidence table. The quantitative results of each study are then usually presented graphically in a “forest plot,” like the one in Box 4 (page 29), which compares the outcomes of industry-funded research with nonindustry-funded research.

If a quantitative synthesis of results is desired, the statistical method of meta-analysis is employed, and a summary result is produced, but this is not always necessary or appropriate. Larger studies that

provide more precise estimates of a treatment's effects are routinely given more weight in the meta-analysis calculations. As shown in Box 4, the length of the line on either side of the "point estimate" (the dots or little black squares) is an indication of the width of the "confidence interval" around the result for each study: the longer the line, the wider the confidence interval—the less precise our estimate of the size of the effect—and the lower our confidence in the estimate. As Box 4 shows, the confidence interval around the summary result (represented by the tail ends of the diamond shape at the bottom of the figure) is very small, meaning that we can have more confidence in it than in the individual studies. That is a key reason for conducting a meta-analysis.

Like systematic reviews, meta-analyses can be of variable quality. A standard method of performing meta-analysis has been developed by the "Quality of Reporting of Meta-Analyses," or QUOROM group. The group published the results of an international conference in the *Lancet* in 1999.

One of the most reliable forms of a systematic review involves collaborating researchers pooling individual patient data from different studies. While not common, this method has been used in a number of cancer-related collaborations, including the early breast cancer collaboration that combined data from 133 trials involving 75,000 women. With such a large amount of data, it is easier to see how particular subgroups of patients fare on the therapy. The precision of the resulting estimates is very high, justifying more confidence in the findings.

INTERPRETING THE REVIEW FINDINGS

The next step in the process of systematic reviewing is returning to the original question that was formulated and explaining how well the results have answered it. Independent peer review of a draft may occur at this point, an important way to help insure the integrity and quality of the final product.

One of the major goals of interpretation is to try to explain the strength of the evidence from the different studies that the review summarized. In other words, for a clinical question about a therapy or procedure, the user of a review needs to know whether the best available evidence comes from study designs at a high level in the hierarchy of evidence—for example, large, well-run randomized controlled trials—or from weaker study designs with less reliable findings. (See Box 1, page 4.) For other types of questions—about the organization or delivery of health services, for example—evidence from randomized controlled trials may not be appropriate. For any type of question, if the only available evidence is expert opinion, some systematic reviews may still summarize it and explain that studies are needed as stronger evidence.

Along with interpreting the strength of the evidence, the systematic review will attempt to assess the quality of the key studies being reviewed, whatever their level of evidence. Randomized controlled trials can be poorly run or well run. A series of individual case studies can be well documented or poorly documented. Providing information about both the level and quality of evidence is a key goal of the systematic review. One tool to help assess the quality of randomized controlled trials is the

BOX 3. WHO DOES GOOD REVIEWS?

Many organizations around the world now produce and disseminate systematic reviews of the evidence, including these:

CANADIAN COORDINATING OFFICE FOR HEALTH TECHNOLOGY ASSESSMENT (CCOHTA)

CCOHTA provides the public with information about health technologies, focusing on evaluations of clinical effectiveness and cost-effectiveness.

THE COCHRANE COLLABORATION

Relying on the work of more than 10,000 people around the world, the collaboration has produced almost 2,000 systematic reviews.

HEALTH EVIDENCE NETWORK (HEN)

Within the World Health Organization, HEN is an information service primarily for public health and health care decision makers in the WHO European region.

INTERNATIONAL NETWORK OF AGENCIES FOR HEALTH TECHNOLOGY ASSESSMENT (INAHTA)

INAHTA facilitates cooperation among 40 agencies from 20 countries, using evidence from systematic reviews as part of its work.

NATIONAL HEALTH SERVICE CENTER FOR REVIEWS AND DISSEMINATION

This group at the University of York in England provides research-based information about the effects of interventions used in health and social care. It conducts systematic reviews as part of its work.

NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE (NICE)

Within the publicly funded National Health Service in England and Wales, NICE provides the public with information on “best practice” and commissions systematic reviews.

SCOTTISH INTERCOLLEGIATE GUIDELINES NETWORK (SIGN)

SIGN produces clinical guidelines based on systematic reviews of the scientific literature. It is part of a new global collaboration called the Guidelines International Network (GIN).

UNITED STATES PREVENTIVE SERVICES TASK FORCE

The task force is an independent panel in primary care and prevention, funded by the U.S. government, that systematically reviews evidence of effectiveness and develops recommendations for clinical preventive services.

EVIDENCE-BASED PRACTICE CENTERS IN THE UNITED STATES AND CANADA

Designated by the Agency for Healthcare Research and Quality (AHRQ) of the U.S. government, these centers all conduct systematic reviews:

Blue Cross and Blue Shield Association, Technology Evaluation Center (TEC),
Chicago, Illinois

Duke University, Durham, North Carolina

ECRI, Plymouth Meeting, Pennsylvania

Johns Hopkins University, Baltimore, Maryland

McMaster University, Hamilton, Ontario, Canada

MetaWorks, Inc., Boston, Massachusetts (1997–2001)

Oregon Health and Science University, Portland, Oregon

RTI International, University of North Carolina at Chapel Hill, North Carolina

Southern California Evidence-Based Practice Center, RAND Corporation,
Santa Monica, California

Stanford University, Stanford, California

Tufts–New England Medical Center (formerly New England Medical Center),
Boston, Massachusetts

University of Alberta, Edmonton, Alberta, Canada

University of California, San Francisco, California

University of Minnesota, Minneapolis, Minnesota

University of Ottawa, Ottawa, Canada

University of Texas Health Science Center, San Antonio, Texas (1997–2001)

Some groups, like ECRI, have very clear policies restricting conflicts of interest in the funding and conduct of reviews. Other groups, including the Cochrane Collaboration, are currently debating the role of sponsors. The source of funding is widely regarded as an important part of the assessment of studies, because sponsor funding is associated with favorable study outcomes.

recently revised CONSORT statement (Consolidated Standards of Reporting Trials) developed to try and improve the design, reporting, and analysis of randomized trials.

Another aspect of interpretation is rigorous assessment of the applicability or generalizability of the results of the systematic review. To what extent is the systematic review relevant and meaningful to a broad or narrow range of different types of patients, drug doses, or policy settings (such as private and public insurance plans)?

Finally, the interpretation may also discuss tradeoffs between benefits and harms, and—less often—costs. Discussions of costs may be called cost-effectiveness analysis, cost-benefit analysis, economic evaluation, or pharmacoeconomics when applied to drugs. The question of whether a treatment or policy intervention—such as sending mothers home early after giving birth—provides value for money in comparison to other options is becoming an increasingly important part of health care decision making, but it is not yet formally integrated within many systematic reviews.

UPDATING THE REVIEW

Updating a review is considered so important by many scientists that it is regarded as the final step in the review process, though getting funding for updates is proving to be a great challenge. The Cochrane Collaboration requires that reviewers consider updating each synthesis every two years. The emergence of important new evidence from a fresh study can mean that updating is needed even sooner. Sometimes the results of a new trial will mean that the updated review will include a new variable, like quality of life or mortality and morbidity measures. Similarly, the updated review may introduce a new method of statistical analysis.

Systematic reviews of evidence are being used in many parts of the health care landscape. People looking for the best evidence to inform their thinking, decision making, and practice include nurses, pharmacists, doctors and other health professionals, patients, insurers, policymakers, advocates, health care executives, and journalists. Similarly, researchers in various academic disciplines are looking at the methods of research synthesis to help assess the current state of knowledge about a particular question, and the gaps in that knowledge, before planning further studies.

High-quality evidence is being presented in different forms for different audiences. According to Andrew Oxman, a former chair of the Steering Group of the Cochrane Collaboration and now director of the Department of Health Services Research at the Norwegian Directorate for Health and Social Welfare, “Systematic reviews are the bricks—the base—needed for building concise, user-friendly health information—whether for doctors, consumers, or anybody else.”

For Jeff Lerner, president and chief executive officer of ECRI (formerly the Emergency Care Research Institute), one of the North American Evidence-Based Practice Centers, everyone involved with health care should know a little about the science of systematic reviews, including the heads of private and public health systems, the executives of major corporations, state insurance commissioners, and elected officials, “and it should be part of the curriculum at every medical school.”

David Eddy, a long-time innovator in evidence-based medical research and currently a senior scientific adviser to the Kaiser Permanente health system, echoes Lerner’s comments but identifies three key groups that should routinely rely on systematic reviews: those writing clinical guidelines; those designing health-system performance measures; and those making health insurance coverage decisions. In this chapter, individuals in different parts of the health care sector explain how they are using systematic reviews.

CLINICIANS

“A high proportion of important clinical questions are simply too complicated for the unaided human mind,” says David Eddy. “So we need a systematic way of addressing those questions and going down to the bedrock of good evidence to get the answers.” After more than a decade of “evidence-based medicine,” a growing number of tools can provide the practicing clinician with accessible, up-to-date summaries of the best evidence about what works and what doesn’t—evidence gathered according to the principles of the systematic review. A good example is *Clinical Evidence*, a textbook written through a transatlantic collaboration involving *BMJ* (formerly the *British Medical Journal*) and the American College of Physicians. Similarly, there are many organizations producing clinical guidelines, including the Scottish Intercollegiate Guidelines Network, that base their recommendations on systematic reviews of the evidence for different health care interventions.

HOSPITALS

“I’m a bit cynical, but I think market forces are more powerful than evidence right now,” says Ian Leverton, vice president of clinical integration with Sutter Health, a not-for-profit group that runs 26 hospitals in California. Leverton says his organization is using evidence, like the reviews produced by ECRI, to help guide decisions about new technologies but notes that formal systems for using the evidence are still being developed. “We are struggling with that. How do we most appropriately disseminate and use this evidence?”

INSURERS

“What we want to make certain is that resources are spent on interventions that we know work, because in medicine the potential for harm is great,” says Allan Korn, chief medical officer of the Blue Cross and Blue Shield Association, which provides systematic review evidence to more than 40 health plans across the United States. According to Korn, the best way to find out if an intervention works is to search the scientific literature systematically. That work is commonly done by the Technology Evaluation Center, an Evidence-Based Practice Center that provides reviews primarily to Blue Cross Blue Shield plans and to Kaiser Permanente. Blue Cross and Blue Shield plans use systematic reviews to inform the deliberations of their medical policy committees, which are responsible for coverage decisions.

Kaiser Permanente also uses systematic reviews to inform some decision making within its insurance arm, in order to help forecast future care costs, and at the service delivery end, to help inform decisions about clinical care. According to Jed Weissberg, associate executive director of quality and performance improvement of the Permanente Federation, Kaiser produces and uses reviews on a daily basis. However, Weissberg sees a need for much better coordination at a national and international level to reduce unnecessary repetition of reviewing. “The whole field needs a great deal of organization and coherence. We’d like to see private and public groups coordinating their efforts. The differences in the methods used by different groups are far less than the similarities.”

JOURNALISTS

“Systematic reviews can help put new findings into the proper context,” says health care journalist Andrew Holtz. “They help reporters and their audiences put things in perspective.” A former CNN health correspondent, Holtz now freelances for a broad range of media organizations and is president of the five-year-old Association of Health Care Journalists. “One of the association’s aims is to help journalists understand the difference between a comprehensive systematic review of the evidence and ... just a literature review that reflects the expert opinions of the authors. A lot of reporters don’t yet know that difference,” says Holtz.

Jessie Gruman, president of the Washington-based Center for the Advancement of Health, agrees that learning more about systematic reviews can improve journalists' reporting on health care. She now runs a service providing journalists with information about studies published in more than 30 journals and has plans to create a new service based upon the results of systematic reviews. "Reading reviews can change the way people cover health news generally, but they can also generate news in their own right."

Holtz says reviews can generate news when they appear to contradict common practice. "If the review says the evidence doesn't support current practice, that's news," he says, citing the controversial review of evidence about mammography screening published in the *Lancet* in 2001. The Association of Health Care Journalists has arranged for its 700 members to receive free access to the Cochrane Library, which includes the full text of Cochrane Reviews. "This means reporters can more quickly get a more accurate picture of what the body of knowledge is for the story they are writing, as opposed to randomly searching recent reports and glancing at abstracts haphazardly. At the very least we can ask the right questions of the researchers," says Holtz. "You can't overestimate the importance of promoting these kinds of reviews for enhancing public understanding of what is known, what's not known, and what yet needs to be researched."

CONSUMERS

"If you want to make an informed decision, this is the way to go," says New York-based consumer advocate and writer Maryann Napoli, explaining the major benefit of using systematic reviews. Napoli works with a group called the Center for Medical Consumers, created almost three decades ago on the premise that people should know more about what works and what doesn't in medicine. "I think the need for systematic reviews of the evidence is a very basic need that I'm always trying to convey to my readers."

Consumer participation has been a key aspect of the development of the international Cochrane Collaboration. For much of the last decade, Australian Hilda Bastian ran the collaboration's international consumer network. Consumers suggest topics for Cochrane Reviews and often work with health care researchers as co-reviewers or referees. While Napoli welcomes this involvement and the availability of Cochrane Reviews to consumers, in the form of free abstracts on the Web, she would prefer that the reviews contained less complicated medical jargon and more user-friendly information. To illustrate, she points to the statistical term "numbers needed to treat" (NNT)—a figure used to demonstrate the magnitude of a treatment's benefit, which is common in evidence-based materials. Napoli says the meaning of NNT is not clear enough to consumers, even to those highly motivated to understand the real magnitude of the risks and benefits of therapies.

POLICYMAKERS

“In making decisions for populations or groups of patients the best available knowledge has to be used and, as with clinical decision making, the systematic review provides the best possible knowledge, although it may not always be available,” wrote Muir Gray in a 2001 textbook on systematic reviews. While many factors other than evidence influence policymakers, and evidence is often used retrospectively to justify a particular course of action, systematic reviews are being used in some policy settings. As program director at the national screening program in the United Kingdom, Gray noted that the government based its decision not to encourage prostate cancer screening upon two systematic reviews of the evidence, neither of which demonstrated any reduction in deaths from screening.

In the United States, the federally funded Medicare program uses systematic reviews of evidence, commonly referred to as evidence reports or technology assessments, to make coverage decisions about medical devices and procedures. Formally announcing updated decision-making rules in 1999, Medicare leaders said the program would now obtain a technology assessment before making any major new national coverage decisions. The *Federal Register* notice that published the rules defined the process of technology assessment: “Generally, a technology assessment provides a systematic analysis of the safety, efficacy, and effectiveness of a health care technology.” That notice also announced more transparency in decision making, with details of coverage decisions to be available on the Web.

A recent case in which a systematic review of the evidence affected a national coverage decision by Medicare concerned a common surgical procedure: arthroscopy for the osteoarthritic knee. Spurred by a trial published in the *New England Journal of Medicine* in 2002, which found no benefit over placebo, Medicare officials systematically reviewed the evidence for the surgery, searching almost 20 databases—including Cochrane Systematic Reviews, the RAND Corporation, ACP Journal Club, etc.—to find all relevant studies and any earlier systematic reviews. They also consulted widely with professional associations. The synthesis of the evidence and summary of their findings, which are available on the Medicare Web site, were unambiguous. Based on the best available evidence, knee arthroscopy for osteoarthritis was not “reasonable and necessary,” and Medicare would therefore not cover certain common forms of the operation.

By way of contrast, following a systematic review of the evidence for cryosurgery for prostate cancer, Medicare officials approved coverage of the procedure as a “primary treatment for clinically localized prostate cancer,” although they noted when they made their decision that it had not been proved effective in all cases of prostate cancer.

For Sean Tunis, chief medical officer of the federal Centers for Medicare & Medicaid Services, systematic reviews help to inform major national policy decisions and serve as an example of credible public policymaking. “The rules of analysis have to be transparent, consistent, and widely accepted. In terms of evaluating the effectiveness of health care interventions, the widely accepted standard is the systematic review. This is how you evaluate clinical effectiveness. You can then offer the public and

stakeholders a process that is predictable and understandable.” Alluding to the many other factors influencing policymaking, Tunis adds, “We live within the boundaries defined by the evidence—but the evidence doesn’t tell us what the policy should be.”

Another example of systematic reviews used by policymakers comes from Australia. Since 1995, federal law has mandated that pharmaceutical companies that want their products listed in Australia’s national formulary must include a systematic review of the available evidence with their submissions. That evidence is then scrutinized and used in the deliberations of the national Pharmaceutical Benefits Advisory Committee, which makes recommendations to the federal government about which new drugs to subsidize and specifies the prices at which those drugs would be more cost-effective than other options, so that the nation receives value for its money.

While Australia’s national system of cost-benefit analysis of new drugs has proved contentious, particularly with the pharmaceutical industry, it is regarded as an innovative international model for evidence-based policymaking. The regulatory requirement for the submission of systematic reviews is partly intended to prevent submission of selective or skewed data, which has occurred in other countries, as with applications for approval of new antidepressants in Sweden.

A more recent example of systematic reviews informing policymaking in the United States is the innovative plan to compare the effectiveness of drugs within each of the 25 most expensive classes. The project grew out of a partnership between the state of Oregon and the Oregon Evidence-Based Practice Center, which subsequently involved other states as well. The current project is being coordinated by the Center for Evidence-Based Policy at the Oregon Health and Science University. It builds on systematic reviews of 12 drug classes that the original partnership produced.

The plan is for the remaining systematic reviews to be conducted and regularly updated by several Evidence-Based Practice Centers across the United States. According to a project overview entitled “Globalizing the Evidence, Localizing the Decisions,” the project will “make available to governments, businesses, non-profit organizations, and citizens the best existing synthesis of global research on the relative effectiveness of similar medications. This research will allow these entities to determine for themselves how best to obtain the most value for their pharmaceutical expenditures.” Project director Mark Gibson, deputy director of the Center for Evidence-Based Policy and who formerly served as a long-time senior health policy official in Oregon, says the plan was originally inspired by the exploding growth in state expenditures for prescription drugs. The top 25 classes account for more than half of all pharmaceutical expenditures by states. (The Milbank Memorial Fund helped to incubate the Center for Evidence-Based Policy.)

One of the most expensive classes was the cholesterol-lowering drugs called statins. “Statins were among the first reviews we did,” says Gibson, referring to the time when he was an adviser to the state’s governor. “The question we were interested in was different from the questions often asked by drug companies. They want to know how well their drugs lowered cholesterol. We wanted to know how their drugs lowered the numbers of heart attacks and strokes.” Oregon conducted that first review of the evidence for five statins in 2002, and according to Gibson the results were astonishing. “At that time

only some of the statins were supported by evidence about the real clinical outcomes, and one was a generic. So the generic went onto our preferred drug list.” For Gibson this was a win-win situation, saving money and shifting physicians’ prescribing patterns to the drugs with the strongest evidence of their benefit. “This is a great object lesson in how far off the mark we can be in terms of the evidence. It’s also a great example of how we can use evidence to make better clinical decisions and better systems decisions.” The plan to produce drug-class reviews has taken off rapidly. So far, 17 states and two nonprofit organizations have agreed to participate and use the results. Gibson expects to have reviews of all 25 classes completed by early 2005.

THE LIMITATIONS OF SYSTEMATIC REVIEWS

Notwithstanding the powerful reasons to base decisions on the strongest evidence from top-quality scientific studies, there are important limitations to the use of systematic reviews. One key limitation is that there is often not enough strong evidence to support or refute a promising treatment or intervention; yet decisions about its use must be made and defended. The methods of synthesizing the results of different studies, both qualitatively and quantitatively, are still evolving and are by no means flawless. Systematic reviews, like all work, can be of varying quality. And no matter how strongly the evidence points in one direction, the values and preferences of health professionals, patients, and policymakers greatly influence health care decisions.

A strong supporter of the move toward evidence-based medicine, with its reliance on systematic reviews, is Australian surgeon and ethicist Miles Little. But he wants to qualify the current enthusiasm for methodological rigor within medicine with a reminder of the centrality of subjective human relationships. As Little points out, evidence-based medicine values the evidence from randomized controlled trials above other evidence, sometimes playing down the importance of other sources of knowledge, such as the personal narratives of those suffering pain and illness. In a recent essay entitled “Better than Numbers,” he observed that health services are “morally based in the intuitive respect we have for human life. . . . [Evidence-based medicine] is the servant of that intuition, not its replacement.”

Marlene Smadu, a former policymaker and now a professor and associate dean of nursing at the University of Saskatchewan in Canada, wonders whether too strong an emphasis on systematic reviews can inhibit our ability to solve some complex health problems. “The tension between asking a question that is focused enough for systematic review,” she says, “yet broad enough that we are not cornered into traditional ways of viewing the world, is real.”

LIMITATIONS OF THE EVIDENCE BEING REVIEWED

The most glaring omission in many studies is the lack of information about potential side effects. According to Andrew Oxman of the Norwegian Directorate for Health and Social Welfare, this is a major problem for most systematic reviews. If the original studies overlook side effects, a review of those studies will also tend to lack this information. Oxman says many trials don’t collect good information on side effects because they concentrate on measuring benefits, and different trials tend to measure different types of side effects, making a summary within a review difficult. Similarly, many trials are too small to pick up rare but potentially serious side effects and too short-lived to pick up long-term side effects. As a result, conscientious reviewers try to use evidence obtained outside of trials, including postmarketing studies and adverse-event reports. But these too often provide inadequate data. “It is extremely difficult to do a good job in this regard,” says Oxman.

A related problem is that drugs attract a much greater share of research funding than other interventions because manufacturers want to develop new products, and regulators are required to

evaluate them. Some observers have even started to ask whether we are moving toward “evidence-based medicine” or “medicine-based evidence.” Research funding for reviews of diagnostic and surgical procedures as well as for complementary approaches to health problems—such as massage therapy, generic drugs, or nonmedical prevention measures—remains very small by comparison.

A further limitation is that information concerning the nature and prevalence of illness, so critical in providing the context for understanding the evidence about the effectiveness of therapies, is sometimes of poor quality. Moreover, the definitions of many conditions are fluid, and there is growing concern that the boundaries defining many “diseases” are being artificially enlarged by those eager to expand markets for medical therapies and services.

A contemporary example of what medical writer Lynn Payer has called “disease-mongering” is occurring with the push to define the complexity of female sexual difficulties as a disease called “female sexual dysfunction,” said by some researchers to affect 43 percent of the female population in the United States. Systematic reviews of evidence about the effects of therapies may have limited value if the definitions of the “diseases” those therapies are targeting are derived more from marketing fiction than from genuine medical need.

THE ROLE OF VESTED INTERESTS

Two systematic reviews published in 2003 concluded that studies funded by industry sponsors are much more likely to find favorable results for the sponsors’ products than studies funded by other sources. (See Box 4, page 29.) Lexchin and colleagues concluded their systematic review in *BMJ* by arguing that there is a “systematic bias” within medical research, because so many original studies are now sponsored by companies and manufacturers with a vested interest in the results. Clearly there is a danger that bias might be passed on when those original studies are included within systematic reviews.

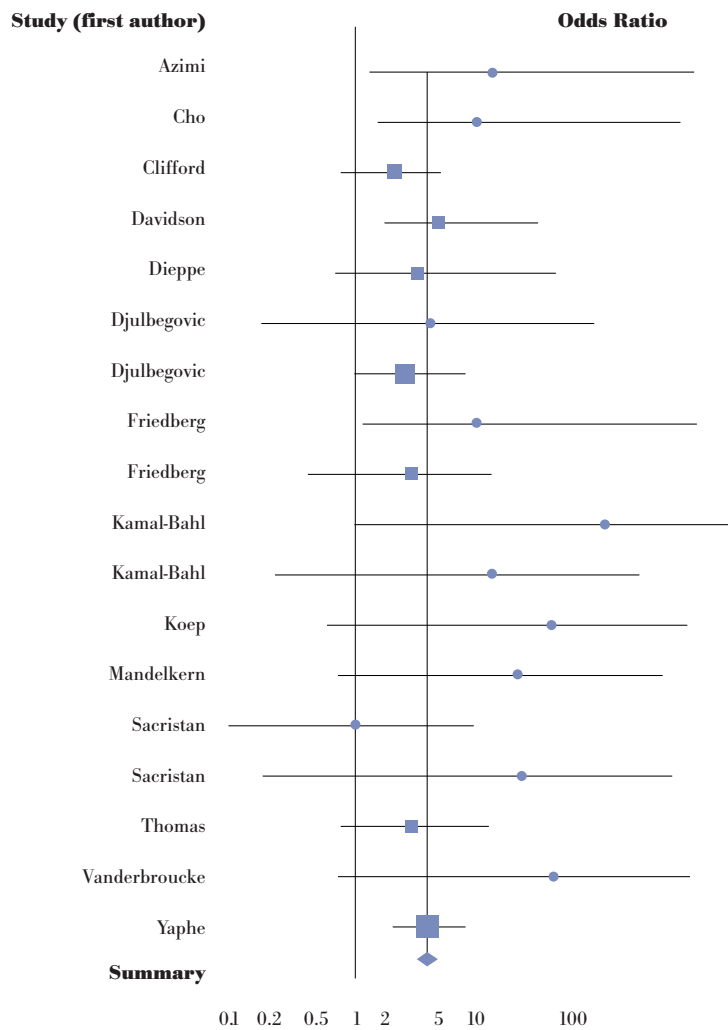
Others concerned about vested interests suspect that those who want to reduce their health care payments, like insurance companies and government agencies, will accord more attention—or will even fund—systematic reviews biased toward unfavorable outcomes, in order to avoid covering expensive but potentially valuable therapies and technologies. Asked about this concern, Jeff Lerner, chief executive of ECRI, acknowledged that private health insurers, for example, could theoretically demand an impossibly high level of proof of the effectiveness of a new technology and then use a review to deny coverage. He was not, however, aware of any study investigating such practices or biases, and he pointed out that one of the publicly funded Evidence-Based Practice Centers regularly producing high-quality systematic reviews is located within the Blue Cross and Blue Shield Association.

Another limitation to consider is the vested interests of the professional groups conducting the original studies or trials that reviewers synthesize. British researcher Nick Black observes that those conducting a systematic review will sometimes “accept the credibility of outcomes in studies without considering biases.” He cites surgical procedures for urinary incontinence as an example. According to Black, in some randomized controlled trials surgeons have reported high rates of success for their

BOX 4. "FOREST PLOT"

This is a "forest plot" showing the results of a systematic review of studies comparing the outcomes of industry-funded research with nonindustry-funded research. The forest plot shows individual odds ratios for the 18 different comparisons (with confidence intervals) and a summary odds ratio at the bottom of the diagram of 4.05 (95% confidence interval 2.98 to 5.51).

This systematic review finds that industry-funded research is four times more likely to have favorable outcomes than research funded by other sources.



Source: Lexchin et al. 2003

procedures. Yet other studies (whose designs are on a lower level in the hierarchy of evidence) find that patients report a much lower success rate for the same procedures. Such discrepancies pose big challenges for those systematically reviewing the evidence.

WHAT HAPPENS WHEN A REVIEW FINDS NO GOOD EVIDENCE?

“For a new treatment, when in doubt hold the line,” is David Eddy’s short answer. “Even when there is a lot of enthusiasm, unless there is good evidence that a new treatment’s benefits outweigh its harms and it provides good value for the money, don’t cover it, don’t identify it as a performance measure, and don’t recommend it in a guideline.” Eddy has been part of the push to bring evidence into decision making in the insurance sector, the movement for quality health care, and in guideline development across a range of different settings—public, private, and professional. A physician and mathematician, he has hands-on experience in all three areas. “I would not forbid the use of an unproven therapy. If you did that, perhaps a third of medical practice would fall away. But the trouble is, with new treatments, when you start doing something, the public and patients quickly think it works, and that can inhibit the incentive and the ability to do the research to find out if it really does work.”

Eddy points to autologous bone marrow transplant (ABMT) for breast cancer as an example of medical practice, and some advocacy groups, getting too far ahead of the evidence. In the late 1980s, published trials with weak study designs that lacked control groups showed promising results for ABMT for breast cancer. Enthusiastic media coverage fed excitement among clinicians and patients, and a number of courts ruled against insurance companies that had denied coverage of the expensive new procedure. By 2000 a number of randomized controlled trials had been completed. They showed that ABMT was ineffective.

Taking another view, others point to the dangers of denying coverage of potentially effective new procedures. As already noted, a lack of evidence of effectiveness is not the same thing as evidence of a lack of effectiveness. New therapies by definition are not supported by strong evidence, so insurers’ refusal to cover them for that reason limits their availability to only the few who can afford to pay their full cost. To get around this problem, Jeff Lerner argues that insurance coverage decisions can be made with the help of a systematic review of the best *available* evidence, however strong it is. And taking a different perspective from Eddy’s, he argues that providing insurance coverage for innovative therapies before strong evidence supports them does not preclude continued rigorous evaluation: “If you do tie coverage decisions to rigorous evaluation, that would deny coverage for vast amounts of care. Our stance is, the patient is in the hospital, and the question is, what is the best evidence available now for their treatments: It may be crappy evidence, it may be fabulous, but that’s the best we can do today.”

Grant Bagley, a partner with the law firm of Arnold and Porter and a former director of the group responsible for insurance coverage decision making within Medicare in the United States, describes the “gray zone” of uncertainty as a genuine challenge for policymakers. “For new technology coming

along, it's the gray zone that is difficult to deal with. If you don't have good evidence, making the transition from not paying to paying in the absence of evidence is the most difficult thing to deal with from a policy point of view." Notwithstanding that difficulty, a good systematic review of whatever evidence is available, even in highly uncertain situations, helps make the policymaking process more transparent and the political or commercial considerations more visible.

THE ROLE OF VALUES

Values, beliefs, and personal preferences play a critical role in health care decision making. As Naomi Aronson, director of the Blue Cross and Blue Shield Association's Technology Evaluation Center, has observed, systematic reviews are not designed to make decisions, simply to assist in making them.

Andrew Oxman cites a recent example where values played a key role in decision making. An international group of researchers reviewed the evidence to support a screening program for melanoma (skin cancer). Like many systematic reviews, this review found only low-level evidence of small benefits for screening, weighed against unknown but plausible harms to patients, including inconvenience, anxiety, and the complications of investigating false positive test results.

Most of the researchers opposed screening, believing that stronger evidence was required before launching an intervention targeting healthy people. Others within the group, with different values, were more comfortable with the uncertainty and believed that screening was likely to do more good than harm. They recommended proceeding with screening. This group placed a "high value on the small but potentially important benefits of screening relative to the unknown potential harms," according to this account of the decision-making process, coauthored by Oxman: "While this example is typical of the value judgments that underlie recommendations about screening, such as PSA [prostatic specific antigen] for prostate cancer, mammography screening for breast cancer and spiral CT [computerized tomography] scans for lung cancer, the same issues arise in making recommendations about therapy for both acute and chronic conditions, where it is always necessary to balance the expected benefits against the expected harms in light of the relative values attached to each important outcome, and uncertainty."

THE CHALLENGES OF META-ANALYSIS

Meta-analysis is a statistical method of combining and summarizing the results from different studies. Some systematic reviews include meta-analysis and some do not. The theory behind meta-analysis is that combining the results of many studies will produce a more accurate and precise result, or set of results, that we can have more confidence in than the results of just one study. It is also a way of understanding why different studies of the same therapy produced different results and can therefore shed light on the particular conditions under which a therapy or intervention will prove most effective. Despite strong scientific support for meta-analysis, critics have pointed to its limitations.

In an article weighing its benefits and limitations in the *Lancet* in 1998, Joseph Lau and colleagues suggested that the results of meta-analysis should be treated with caution but concluded, on balance, that it was a method of synthesis worth pursuing. The article's authors also argued that it might sometimes be better to come up with more than one result from meta-analysis, as a single summary finding could be misleading: "Meta-analysis may be better suited to explore differences in harm and benefit across subgroups of patients."

One of the most strident criticisms of meta-analysis comes from H.J. Eysenck, who has expressed a range of concerns, most notably that the method often combines too many different kinds of studies, like adding apples and oranges. He also argues that those conducting meta-analysis need to be more sensitive to clinical reality, in order to avoid including treatments that are no longer relevant to actual practice.

A common rationale offered by proponents of meta-analysis is that pooling the results of several studies sometimes reveals a beneficial treatment effect that could not be detected in the smaller studies. An example often cited is the use of steroids by women about to give birth to premature babies. Several small studies had failed to detect any favorable effects of this treatment. However, when the results from the individual studies were combined, the total number of patients was big enough to reveal a statistically and clinically significant benefit. The results of that meta-analysis have changed the care of women delivering premature babies.

Eysenck, however, is skeptical of that kind of rationale, writing that "if a medical treatment has an effect so recondite and obscure as to require meta-analysis to establish it, I would not be happy to have it used on me. It would seem better to improve the treatment, and the theory underlying the treatment." However, defenders of meta-analysis point out that many medical interventions have only small to moderate effects.

Another critique comes from John Bailar of the University of Chicago, who wrote a commentary in the *New England Journal of Medicine* in 1997 entitled "The Promise and Problems of Meta-Analysis." In the same journal issue, a group of researchers published a study comparing the results of a series of meta-analyses with the results of large, individual, randomized controlled trials about the same therapies. The results of several of the meta-analyses seemed to overstate the benefits of treatments when compared to the results of the individual trials. Pointing to these discrepancies, Bailar argues that meta-analysis is not a sure way of distilling "the truth" and rejects the idea that a summary score is the best estimate of how well a treatment works. "I still prefer conventional narrative reviews of the literature."

Others echo Bailar's opinions by suggesting that systematic reviews sometimes lack a thoughtful understanding of the context for the interventions being reviewed and maintaining that informed creative judgment as well as methodological rigor is required to synthesize a body of knowledge. For example, Howard Goldman, professor of psychiatry at the University of Maryland, says that "evidence reviews... sometimes miss... important policy nuances. For example, while the Cochrane Review of integrated treatment for co-occurring severe mental illness and substance misuse provided a comparatively negative assessment of the intervention, based on a review of the same evidence the

mental health report of the Surgeon General considered integrated treatment as an effective intervention. That review concluded that it was the only intervention demonstrated effective for a very important public health problem, and in clinical trials it was always superior to treatment as usual.”

One of the architects of evidence-based health care, Gordon Guyatt of McMaster University, acknowledges limitations to the use of meta-analysis but takes issue with some of its critics. The biggest problem, he agrees, comes from the attempt to combine the results of studies with apparently very different findings, referred to as a situation where there is heterogeneity. In these cases the challenge is to try to understand and explain the heterogeneity, says Guyatt, rather than to abandon attempts to synthesize the results: “I start from the point of how we make clinical decisions for our patients. We should get our best estimates of the benefits and risks of treatment and weigh them up, ideally with patients. My question to critics of systematic reviews and meta-analysis is, ‘If we are not going to use these approaches, how are we going to get these best estimates?’ With all the limitations of meta-analysis, I don’t see a better way.”

There is growing support around the world for producing and using systematic reviews of health care interventions. Notwithstanding their limitations, systematic reviews are changing the way decisions are made within health care and beyond. The newly formed Campbell Collaboration is now producing systematic reviews of evidence about the effects of interventions in social welfare, education, and criminal justice.

Several registers have been set up in different parts of the world to document all clinical trials, with the eventual goal that all their results will be made public, whether positive or negative. This should significantly improve the quality of systematic reviews by combating the problem of publication bias, whereby negative findings tend to remain hidden from view. Several pharmaceutical companies have joined this effort to varying degrees. It is hoped that those reviewing original studies in the future can get access to all the data, not just the part that researchers or sponsors wish to reveal.

There are encouraging signs that the quality of systematic reviews is improving in other ways. The methods of performing and applying meta-analysis are steadily improving, according to a study published in 1996. At the same time, there is growing interest in big reviews that can pool and analyze individual patient data from many trials, like the international collaboration on the treatment of early breast cancer. By combining such large amounts of data, researchers developed much more reliable and precise estimates of how well breast cancer drugs and other therapies work for particular kinds of women.

Updating the existing systematic reviews is a challenge for all those who produce them, but the growing recognition and kudos accorded to them by the academy, medical journals, and public and private insurers seem to ensure a continued supply of willing reviewers. The estimated median length of time needed to complete a systematic review is equivalent to 30 weeks of full-time work by one person. Though the length of time can vary dramatically, depending on the complexity of the systematic review, there is clearly a big potential scientific return for a modest investment. In the case of a commonly used, expensive, unproven operation, a review of the evidence that could tell us how well that procedure works might cost less than just one operation.

Growing recognition of the value of systematic reviews will probably increase their use by researchers and journalists writing about health care. Previous studies have found that authors of journal articles often cite references selectively to support their views or findings and that few contextualize new findings within the previous body of knowledge—as represented by a systematic review of what is known. While many health care journalists remain unfamiliar with the science of systematically reviewing, that may soon change: The Association for Health Care Journalists is offering its members free access to the full text of Cochrane Reviews and holding workshops on some of the basic principles of evidence-based health care.

Many challenges remain. Identifying and minimizing the influence of vested interests in both original studies and systematic reviews is an enormous task. Improved methods for synthesizing existing systematic reviews—that is, synthesizing the syntheses—will be needed as the number of reviews grows. Rigorous methods for assessing evidence to guide decisions about whether new

interventions offer value for money require a lot of attention. Perhaps most fundamentally, it will take considerable work to make sure that the best evidence is applied *in practice*.

In many policymaking forums, recognition of the value of systematically reviewing evidence appears to be increasing. In the United Kingdom, the National Institute for Clinical Excellence and other national and regional bodies routinely commission systematic reviews. In the United States and Canada, Evidence-Based Practice Centers are producing high-quality evidence reports and technology assessments for various purchasers. Innovative projects, like the drug class reviews being developed by Oregon's Institute for Evidence-Based Policy, are emerging so that the new science of research synthesis can meet the needs of policymakers.

Even when those making decisions choose to overlook the evidence, the availability of a systematic review can bring greater transparency to the decision-making process. Moreover, the potential benefits flow in several directions. The growing demand for systematic reviews can help lift the quality of health care research and make it more relevant and useful to health care professionals, policymakers, and the wider community.

Finally, though, no matter how good the evidence is, evidence doesn't make decisions. People do. But perhaps with the evidence in hand, some future decisions might be just a little more informed.

NOTES

EXECUTIVE SUMMARY

p.1 *As Richard Light and David Pillemer explain:* Light and Pillemer 1984, viii.

A NOTE ABOUT NOMENCLATURE

- 2 *The most widely used term for the process is:* Based on definition of Cochrane Collaboration at <http://www.cochrane.org/resources/glossary.htm> (accessed Nov. 20, 2003).
- 2 *“Do antidepressants help smoking cessation:* <http://www.cochrane.org/cochrane/revabstr/AB000031.htm> (accessed Nov. 24, 2003).
- 2 *Internationally the Cochrane Collaboration produces:* <http://www.cochrane.org> (accessed Nov. 24, 2003).
- 2 *In the United States and Canada:* <http://www.ahrq.gov/clinic/epcix.htm> (accessed Nov. 24, 2003).
- 2 *The International Network of Agencies:* <http://www.inahta.org> (accessed Nov. 13, 2003).
- 2 *The World Health Organization’s Health Evidence Network produces:* <http://www.who.dk/eprise/main/WHO/Progs/HEN/Home> (accessed Nov. 13, 2003).

INTRODUCTION TO SYSTEMATIC REVIEWS

- 4 *Several such hierarchies have been produced:* Upshur 2003, 672–3.
- 4 *The International GRADE Working Group is:* Schünemann et al. 2003, 677–80.
- 4 *A simple hierarchy, from the strongest level to the weakest:* http://www.cebm.net/levels_of_evidence.asp (accessed Nov. 20, 2003).
- 5 *It found that quitting smoking:* Critchley and Capewell 2003.
- 5 *When rigorous scientific evaluation does:* Rossouw et al. 2002.

THE HISTORY OF SYSTEMATIC REVIEWS

- 6 *A perfect introduction to this dual history comes:* Lind 1753.
- 6 *On board the Salisbury ... Lind took:* <http://www.jameslindlibrary.org/> (accessed Nov. 24, 2003).
- 6 *Six years later, the results were published:* Ibid.
- 6 *In 1904 Karl Pearson ... published:* Chalmers, Hedges, and Cooper 2002.
- 6 *Three years later a scientist in the United States reviewed:* Ibid.
- 7 *In 1971 Archie Cochrane ... published:* Cochrane 1972.
- 7 *In 1984 Richard Light and David Pillemer published:* Light and Pillemer 1984.
- 7 *Three years later the science of research synthesis leapt:* Mulrow 1987.
- 7 *In 1993 Oxman and Guyatt published:* Oxman and Guyatt 1993.
- 8 *One milestone was a 1979 report:* Canadian Task Force on the Periodic Health Examination 1979.

- 8 *Reviewing the evidence regarding prostate surgery:* Wennberg et al. 1993.
- 8 *The PORT examining total knee replacement searched:* Freund, Katz, and Callahan 1993.
- 8 *The PORT concerned with treatments for chronic lower-back pain concluded:* Deyo et al. 1997.
- 8 *As Iain Chalmers ... and his colleagues report:* Chalmers, Hedges, and Cooper 2002.
- 8 *The resulting publication summarized:* Enkin, Keirse, and Chalmers 1989.
- 9 *Its electronic database, the Cochrane Library, now boasts:* Moynihan 2003b.
- 9 *Abstracts of those reviews are:* <http://www.cochrane.org> (accessed Nov. 24, 2003).
- 9 *In 1995 Australia's federal government mandated:* Hill, Henry, and Stevens 2001.
- 9 *A number of national research funding agencies ... began:* Chalmers, Hedges, and Cooper 2002.
- 9 *The national research ethics committee in Denmark went:* *Ibid.*, 27.
- 9 *These centers, which now number 13, conduct:* <http://www.ahrq.gov/about/profile.htm> (accessed Nov. 24, 2003).
- 9 *In 1999 the U.S. agency responsible for Medicare announced:* <http://www.cms.hhs.gov/coverage> (accessed Nov. 24, 2003).

WHY DO SYSTEMATIC REVIEWS?

- 10 *As Light and Pillemer explain:* Light and Pillemer 1984, viii.
- 10 *Cynthia Mulrow ... argues:* Mulrow 1995.
- 10 *In 1998 Mulrow cowrote:* Mulrow and Cook 1998, 3.
- 10 *There are now millions of studies:* Mulrow 1995, 2.
- 10 *Nicholas Mays ... recently outlined:* Mays, Roberts, and Popay 2001, 190.
- 11 *Light and Pillemer have written:* Light and Pillemer 1984, 9.
- 11 *In July 1992, the Journal of the American Medical Association (JAMA) published:* Antman et al. 1992.
- 11 *Nonrandomized trials of anticoagulant medicines exaggerated:* Chalmers et al. 1977.
- 11 *Nonblinded studies of low-molecular-weight heparin exaggerated:* Juni et al. 1999.
- 11 *Nonblinded studies of pregnancy and childbirth therapies exaggerated:* Schulz et al. 1995.
- 11 *Poorly randomized trials of pregnancy and childbirth therapies exaggerated:* *Ibid.*
- 11 *Low-quality studies of interventions ... exaggerated:* Moher et al. 1998.
- 11 *A systematic review of all the relevant evidence suggested:* Kunz, Vist, and Oxman 2004.
- 12 *Using a statistical method called:* Antman et al. 1992.
- 12 *Thomas Moore ... has shown that widespread use:* Moore 1995.
- 12 *While a systematic review of the evidence suggested:* Ferburg 1983.
- 12 *"The warning signs were there:* Chalmers 2001.
- 13 *One of the earliest individual trials of antiarrhythmic drugs found:* Cowley et al. 1993; Egger, Dickersin, and Davey Smith 2001.
- 13 *Antman and colleagues summarized:* Antman et al. 1992.
- 13 *As president of the National Business Group on Health:* Author's interview with Helen Darling, 2003.

HOW TO DO A SYSTEMATIC REVIEW

- 14 *Different manuals enumerate different steps:* Glasziou et al. 2001; Khan et al. 2003; Mulrow and Cook 1998.
- 14 *As the Practical Guide explains:* Glasziou et al. 2001.
- 15 *As the reviewers' handbook ... points out:* <http://www.cochrane.dk/cochrane/handbook/hbook.htm> (accessed Nov. 24, 2003).
- 15 *The details of search methods are explained:* see Glasziou et al. 2001; Khan et al. 2003; Mulrow and Cook 1998.
- 15 *Electronic databases such as CENTRAL, EMBASE, and MEDLINE:* <http://www.cochrane.org.us/centralmn.htm#SC> (accessed Feb. 25, 2004); <http://www.embase.com> (accessed Feb. 26, 2004); <http://www.pubmed.gov> (accessed Feb. 26, 2004).
- 15 *If reviewers are searching for studies:* Mays, Roberts, and Popay 2001, 189.
- 15 *Because studies that find no effect for a treatment or intervention are:* Egger, Dickersin, and Davey Smith 2001.
- 15 *In 1986 the Australian researcher John Simes demonstrated:* Simes 1986.
- 15 *Simes's findings have been echoed:* Egger, Dickersin, and Davey Smith 2001.
- 16 *According to Glasziou and colleagues, there are two parts:* Glasziou et al. 2001.
- 16 *There are many different scales:* West et al. 2002.
- 16 *The resulting report summarized the findings:* Ibid.
- 16 *For example, one scale deemed to be valuable was:* Barnes and Bero 1998.
- 16 *In 1999 Peter Juni and colleagues took:* Juni et al. 1999.
- 16 *Similarly, part of the controversy ... arose:* Olsen and Gøtzsche 2001.
- 16 *Disagreements over approaches are likely to persist:* Schünemann et al. 2003.
- 17 *Like systematic reviews, meta-analyses can be:* Moher et al. 1999.
- 17 *While not common, this method has been used:* Early Breast Cancer Trialists' Collaborative Group 1992.
- 17 *One tool to help assess the quality:* Moher et al. 2003.
- 18 *Canadian Coordinating Office for Health Technology Assessment:* <http://www.ccohta.ca> (accessed Nov. 13, 2003).
- 18 *The Cochrane Collaboration:* <http://www.cochrane.org> (accessed Nov. 24, 2003).
- 18 *Health Evidence Network:* <http://www.who.dk/eprise/main/WHO/Progs/HEN/Home> (accessed Nov. 13, 2003).
- 18 *International Network of Agencies for Health Technology Assessment:* <http://www.inahta.org/> (accessed Nov. 13, 2003).
- 18 *National Health Service Center for Reviews and Dissemination:* <http://www.york.ac.uk/inst/crd/> (accessed Nov. 24, 2003).
- 18 *National Institute for Clinical Excellence:* <http://www.nice.org.uk/cat.asp?c=137> (accessed Nov. 24, 2003).

- 18 *Scottish Intercollegiate Guidelines Network*: <http://www.sign.ac.uk/> (accessed Nov. 24, 2003).
- 18 *SIGN produces clinical guidelines*: <http://www.ahrq.gov/clinic/uspstfix.htm> (accessed Nov. 24, 2003).
- 18 *United States Preventive Services Task Force*: <http://www.g-i-n.net/index.cfm?fuseaction=links> (accessed Nov. 24, 2003).
- 19 *Evidence-Based Practice Centers in the United States and Canada*: <http://www.ahrq.gov/clinic/epcix.htm> (accessed Nov. 24, 2003).
- 19 *Blue Cross and Blue Shield Association*: <http://www.bcbs.com/tec/index.html> (accessed Nov. 24, 2003).
- 19 *Duke University*: <http://www.clinpol.mc.duke.edu/ProjectDir/EvidenceBasedMedicine/EPC/epc.html> (accessed Nov. 24, 2003).
- 19 *ECRI, Plymouth Meeting*: http://www.ecri.org/about/AboutECRI_Frameset.asp (accessed Nov. 24, 2003).
- 19 *Johns Hopkins University*: <http://www.hopkinsmedicine.org/epc/> (accessed Nov. 24, 2003).
- 19 *McMaster University*: <http://hiru.mcmaster.ca/epc/> (accessed Nov. 24, 2003).
- 19 *MetaWorks, Inc.*: <http://www.metaworks.com> (accessed Nov. 24, 2003).
- 19 *Oregon Health and Science University*: <http://www.ohsu.edu/epc> (accessed Nov. 24, 2003).
- 19 *RTI International, University of North Carolina at Chapel Hill*: <http://www.rti.org/epc/home.html> (accessed Nov. 24, 2003).
- 19 *The source of funding is widely regarded*: Lexchin et al. 2003.
- 20 *Sometimes the results of a new trial will mean*: <http://www.cochrane.dk/cochrane/handbook/hbook.htm> (accessed Nov. 24, 2003).

THE USES OF SYSTEMATIC REVIEWS

- 21 *According to Andrew Oxman*: Author's interview with Andrew Oxman, 2003.
- 21 *For Jeff Lerner, president and chief executive officer of ECRI*: Author's interview with Jeff Lerner, 2003.
- 21 *David Eddy, a long-time innovator in evidence-based medical research*: Author's interview with David Eddy, 2003.
- 21 *"A high proportion of important clinical questions are*: Ibid.
- 21 *A good example is Clinical Evidence*: <http://www.clinicalevidence.com/lpBinCE/lpext.dll?f=templates&fn=main-h.htm&2.0> (accessed Nov. 24, 2003).
- 21 *Similarly, there are many organizations*: <http://www.sign.ac.uk/> (accessed Nov. 24, 2003).
- 22 *"I'm a bit cynical*: Author's interview with Ian Leverton, 2003.
- 22 *"What we want to make certain is*: Author's interview with Allan Korn, 2003.
- 22 *According to Jed Weissberg*: Author's interview with Jed Weissberg, 2003.
- 22 *"Systematic reviews can help*: Author's interview with Andrew Holtz, 2003.
- 23 *Jessie Gruman ... agrees*: Author's interview with Jessie Gruman, 2003.
- 23 *Holtz says reviews can generate news*: Author's interview with Andrew Holtz, 2003; Olsen and Götzsche 2001.

- 23 *“If you want to make an informed decision: Author’s interview with Maryann Napoli, 2003.*
- 23 *While Napoli welcomes this involvement: Ibid.*
- 24 *“In making decisions for populations: Gray 2001, 412.*
- 24 *As program director at the national screening program in the United Kingdom: Gray 2001.*
- 24 *Formally announcing updated decision-making rules in 1999: <http://www.cms.hhs.gov/coverage/> (accessed Nov. 24, 2003).*
- 24 *The Federal Register notice that published the rules: http://www.cms.hhs.gov/mcd/index_list.asp?list_type=ncd (accessed Nov. 24, 2003).*
- 24 *A recent case in which a systematic review of the evidence affected: <http://www.cms.hhs.gov/ncdr/memo.asp?id=7> (accessed Nov. 24, 2003).*
- 24 *Spurred by a trial published: Moseley et al. 2002.*
- 24 *By way of contrast . . . Medicare officials approved: <http://www.cms.hhs.gov/ncdr/memo.asp?id=81> (accessed Nov. 26, 2003).*
- 24 *For Sean Tunis . . . systematic reviews help to inform: Author’s interview with Sean Tunis, 2003.*
- 25 *Since 1995, federal law has mandated: Hill, Henry, and Stevens 2001.*
- 25 *While Australia’s national system of cost-benefit analysis of new drugs has proved: Maynard et al. 2002.*
- 25 *The regulatory requirement for the submission of systematic reviews is: Melander et al. 2003.*
- 25 *According to a project overview entitled: Gibson 2003.*
- 25 *Project director Mark Gibson . . . says the plan: Author’s interview with Mark Gibson, 2003.*
- 25 *“Statins were among the first reviews we did: Ibid.; Drug Class Review on HMG-CoA Reductase Inhibitors (Statins) 2003.*

THE LIMITATIONS OF SYSTEMATIC REVIEWS

- 27 *Systematic reviews, like all work, can be: Shea et al. 2002.*
- 27 *A strong supporter of the move toward evidence-based medicine: Little 2003.*
- 27 *Marlene Smadu . . . wonders whether: Personal communication with the author.*
- 27 *According to Andrew Oxman . . . this is: Author’s interview with Andrew Oxman, 2003.*
- 28 *Some observers have even started to ask: Dieppe 1998.*
- 28 *Moreover, the definitions of many conditions are fluid: Moynihan 2003a; Moynihan, Heath, and Henry 2002; Payer 1990, 1992; Woloshin et al. 2001.*
- 28 *A contemporary example of what medical writer Lynn Payer has called: Laumann, Paik, and Rosen 1999; Moynihan 2003c.*
- 28 *Two systematic reviews published in 2003 concluded: Bekelman, Li, and Gross 2003; Lexchin et al. 2003.*
- 28 *Asked about this concern, Jeff Lerner . . . acknowledged: Author’s interview with Jeff Lerner, 2003.*
- 28 *British researcher Nick Black observes: Personal communication with Nick Black, 2003.*
- 29 *This is a “forest plot” showing the results: Lexchin et al. 2003.*
- 30 *“For a new treatment, when in doubt: Author’s interview with David Eddy, 2003.*

- 30 *Eddy points to autologous bone marrow transplant*: Welch and Mogielnicki 2002.
- 30 *To get around this problem, Jeff Lerner argues*: Author's interview with Jeff Lerner, 2003.
- 30 *Grant Bagley, a partner with the law firm*: Author's interview with Grant Bagley, 2003.
- 31 *Most of the researchers opposed screening*: Personal communication with Andrew Oxman, 2003.
- 32 *In an article weighing its benefits*: Lau, Ioannidis, and Schmid 1998.
- 32 *One of the most strident criticisms of meta-analysis comes*: Eysenck 1995.
- 32 *However, when the results from the individual studies were combined*: <http://www.cochrane.org/cochrane/cc-broch.htm#LOGO> (accessed Nov. 26, 2003).
- 32 *Eysenck, however, is skeptical of that kind of rationale*: Eysenck 1995, 73.
- 32 *Another critique comes from John Bailar*: Bailar 1997.
- 32 *In the same journal issue, a group of researchers published*: LeLorier et al. 1997.
- 32 *For example, Howard Goldman, professor of psychiatry*: Howard Goldman to Andrew Gyory, e-mail, Dec. 15, 2003; U.S. Department of Health and Human Services 1999.
- 33 *One of the architects of evidence-based health care*: Author's interview with Gordon Guyatt, 2003.

THE FUTURE OF SYSTEMATIC REVIEWS

- 34 *The newly formed Campbell Collaboration is*: <http://www.campbellcollaboration.org/> (accessed Nov. 26, 2003).
- 34 *Several registers have been set up*: Dickersin and Rennie 2003.
- 34 *The methods of performing and applying meta-analysis are*: Sacks et al. 1996.
- 34 *At the same time, there is growing interest in*: Early Breast Cancer Trialists' Collaborative Group 1992.
- 34 *Updating the existing systematic reviews is*: Based on the author's interview with Mike Clarke, co-chair of Cochrane Collaboration Steering Group, 2003.
- 34 *The estimated median length of time needed to complete*: Glasziou et al. 2001.
- 34 *Previous studies have found*: Gotzsche 1987.
- 34 *And that few contextualize new findings within the previous body of knowledge*: Clarke and Chalmers 1998.
- 35 *In the United Kingdom, the National Institute for Clinical Excellence*: Author's interview with Ron Stamp, 2003.
- 35 *Innovative projects, like the drug class reviews being developed*: Author's interview with Andrew Oxman, 2003.

REFERENCES

- Antman, E., J. Lau, B. Kupelnick, F. Mosteller, and T. Chalmers. 1992. A Comparison of Results of Meta-Analyses of Randomized Control Trials and Recommendations of Clinical Experts. *Journal of the American Medical Association* 268:240–8.
- Bailar, J. 1997. The Promise and Problems of Meta-Analysis. *New England Journal of Medicine* 337:559–60.
- Barnes, D., and L. Bero. 1998. Why Review Articles on the Health Effects of Passive Smoking Reach Different Conclusions. *Journal of the American Medical Association* 279:1566–70.
- Bekelman, J., Y. Li, and C.P. Gross. 2003. Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review. *Journal of the American Medical Association* 289:454–65.
- Canadian Task Force on the Periodic Health Examination. 1979. Task Force Report: The Periodic Health Examination. *Canadian Medical Association Journal* 121:1193–1254.
- Chalmers, I. 2001. Using Systematic Reviews and Registers of Ongoing Trials for Scientific and Ethical Trial Design, Monitoring, and Reporting. In *Systematic Reviews in Health Care: Meta-Analysis in Context*, 2nd ed., edited by M. Egger, G. Davey Smith, and D. Altman, 429–46. London: BMJ Books.
- Chalmers, I., L. Hedges, and H. Cooper. 2002. A Brief History of Research Synthesis. *Evaluation and the Health Professions* 25:12–37.
- Chalmers, T.C., R.J. Matta, H. Smith, and A. Kunzler. 1977. Evidence Favoring the Use of Anticoagulants in the Hospital Phase of Acute Myocardial Infarction. *New England Journal of Medicine* 297:1091–6.
- Clarke, M., and I. Chalmers. 1998. Discussion Sections in Reports of Controlled Trials Published in General Medical Journals: Islands in Search of Continents. *Journal of the American Medical Association* 280:280–2.
- Cochrane, A.L. 1972. *Effectiveness and Efficiency. Random Reflections on Health Services. The Nuffield Provincial Hospitals Trust*. Cambridge, England: Cambridge University Press.
- Cowley, A.J., A. Skene, K. Stainer, and J.R. Hampton. 1993. The Effect of Lorcaïnide on Arrhythmias and Survival in Patients with Acute Myocardial Infarction: An Example of Publication Bias. *International Journal of Cardiology* 40:161–6.

- Critchley, J., and S. Capewell. 2003. Smoking Cessation for the Secondary Prevention of Coronary Heart Disease (Cochrane Methodology Review). In *Cochrane Library*, Issue 4. Chichester, England: John Wiley and Sons. Available at <http://www.cochrane.org/cochrane/revabstr/ab00304l.htm> (accessed Dec. 7, 2003).
- Deyo, R., B. Psaty, G. Simon, E. Wagner, and G. Omenn. 1997. The Messenger Under Attack—Intimidation of Researchers by Special-Interest Groups. *New England Journal of Medicine* 336:1176–80.
- Dickersin, K., and D. Rennie. 2003. Registering Clinical Trials. *Journal of the American Medical Association* 290:516–23.
- Dieppe, P. 1998. Evidence-Based Medicine or Medicine-Based Evidence? *Annals of the Rheumatic Disease* 57:385–6.
- Drug Class Review on HMG-CoA Reductase Inhibitors (Statins). 2003. Oregon Health and Science University Final Report. Available at <http://www.oregonrx.org/OrgrxPDF/Statins%20review/EPC%20Reports/STATIN%20Update%20Report/statins%20July%202003.pdf> (accessed Nov. 24, 2003).
- Early Breast Cancer Trialists' Collaborative Group. 1992. Systematic Treatment of Early Breast Cancer by Hormonal, Cytotoxic, or Immune Therapy: 133 Randomised Trials Involving 31,000 Recurrences and 24,000 Deaths among 75,000 Women. *Lancet* 339:1–15, 71–85.
- Egger, M., K. Dickersin, and G. Davey Smith. 2001. Problems and Limitations in Conducting Systematic Reviews. In *Systematic Reviews in Health Care: Meta-Analysis in Context*, 2nd ed., edited by M. Egger, G. Davey Smith, and D. Altman, 43–68. London: BMJ Books.
- Enkin, M., M. Keirse, and I. Chalmers. 1989. *A Guide to Effective Care in Pregnancy and Childbirth*. Oxford, England: Oxford University Press.
- Eysenck, H.J. 1995. Problems with Meta-Analysis. In *Systematic Reviews*, edited by I. Chalmers and D. Altman, 64–74. London: BMJ Books.
- Ferburg, C.D. 1983. Effect of Antiarrhythmic Drugs on Mortality after Myocardial Infarction. *American Journal of Cardiology* 52:32C–36C.

- Freund, D., B. Katz, and C. Callahan. 1993. Patient Outcomes Research Teams: Examples from a Study on Knee Replacement. In *Doing More Good than Harm: The Evaluation of Health Care Interventions*. New York: Annals of the New York Academy of Sciences 703:86–95.
- Gibson, M. 2003. Globalizing the Evidence, Localizing the Decisions. An informal working document made available to the author.
- Glasziou, P., L. Irwig, C. Bain, and G. Colditz. 2001. *Systematic Reviews in Health Care: A Practical Guide*. Cambridge, England: Cambridge University Press.
- Gotzsche, P.C. 1987. Reference Bias in Reports of Drug Trials. *BMJ* 295:654–6.
- Gray, M. 2001. Using Systematic Reviews for Evidence Based Policy Making. In *Systematic Reviews in Health Care: Meta-Analysis in Context*, 2nd ed., edited by M. Egger, G. Davey Smith, and D. Altman, 410–8. London: BMJ Books.
- Hill, S., D. Henry, and A. Stevens. 2001. The Use of Evidence in Drug Selection: The Australian Pharmaceutical Scheme. In *Informing Judgment: Case Studies of Health Policy and Research in Six Countries*, 7–27. New York: Milbank Memorial Fund and Cochrane Collaboration.
- Juni, P., A. Witschi, R. Bloch, and M. Egger. 1999. The Hazards of Scoring the Quality of Clinical Trials for Meta-Analysis. *Journal of the American Medical Association* 282:1054–60.
- Khan, K., R. Kunz, J. Kleijnen, and G. Antes. 2003. *Systematic Reviews to Support Evidence-based Medicine*. London: Royal Society of Medicine Press.
- Kunz, R., G. Vist, and A.D. Oxman. 2004. Randomisation to Protect against Selection Bias in Healthcare Trials (Cochrane Methodology Review). *Cochrane Library* 1. Chichester, U.K.: John Wiley and Sons.
- Lau, J., J. Ioannidis, and C. Schmid. 1998. Summing Up Evidence: One Answer Is Not Always Enough. *Lancet* 351:123–7.
- Laumann, E., A. Paik, and R. Rosen. 1999. Sexual Dysfunction in the United States. *Journal of the American Medical Association* 281:537–44.
- LeLorier, J., G. Grégoire, A. Benhaddad, J. Lapierre, and F. Derderian. 1997. Discrepancies between Meta-Analyses and Subsequent Large Randomized, Controlled Trials. *New England Journal of Medicine* 337:536–42.

Lexchin, J., L. Bero, B. Djulbegovic, and O. Clark. 2003. Pharmaceutical Industry Sponsorship and Research Outcome and Quality: Systematic Review. *BMJ* 326:1167–70.

Light R., and D. Pillemer. 1984. *Summing Up: The Science of Reviewing Research*. Cambridge, Mass.: Harvard University Press.

Lind, J. 1753. *A Treatise of the Scurvy in Three Parts. Containing an inquiry into the Nature, Causes and Cure of that Disease, together with a Critical and Chronological View of what has been published on the subject*. London: A. Millar. Available at [http://pc-78-120.udac.se:8001/WWW/Nautica/Medicine/Lind\(1753\).html](http://pc-78-120.udac.se:8001/WWW/Nautica/Medicine/Lind(1753).html) (accessed Nov. 24, 2003).

Little, M. 2003. Better than Numbers...: A Gentle Critique of Evidence-Based Medicine. *Australian and New Zealand Journal of Surgery* 73:177–82.

Maynard, A., R. Cookson, D. McDaid, F. Sassi, T. Sheldon, and the ASTEC Group. 2002. Analysis of the Scientific and Technical Evaluation of Health Interventions in the European Union (A Report from the London School of Economics). Available at http://www.lse.ac.uk/Depts/lsehsc/pdf_files/outside.pdf (accessed Nov. 24, 2003).

Mays, N., E. Roberts, and J. Popay. 2001. Synthesising Research Evidence. In *Studying the Organisation and Delivery of Health Services*, edited by N. Fulop, P. Allen, A. Clarke, and N. Black, 188–220. London: Routledge.

Melander, H., J. Ahlqvist-Rastad, G. Meijer, and B. Beermann. 2003. Evidence B(i)ased Medicine—Selective Reporting from Studies Sponsored by Pharmaceutical Industry: Review of Studies in New Drug Applications. *BMJ* 326:1171–3.

Moher, D., D. Cook, S. Eastwood, I. Olkin, D. Rennie, and D. Struop for the Quorum Group. 1999. Improving the Quality of Reports of Meta-Analyses of Randomised Controlled Trials: The QUOROM Statement. Quality of Reporting of Meta-Analyses. *Lancet* 354:1896–1900.

Moher, D., B. Pham, A. Jones, D. Cook, A. Jadad, M. Moher, and P. Tugwell. 1998. Does Quality of Reports of Randomised Trials Affect Estimates of Intervention Efficacy Reported in Meta-Analyses? *Lancet* 352:609–13.

Moher, D., K.F. Shulz, D.G. Altman, and the CONSORT Group. 2003. The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomised Trials. *Clinical Oral Investigations* 7:2–7.

- Moore, T., 1995. *Deadly Medicine*. New York: Simon and Schuster.
- Moseley, J.B., K. O'Malley, N.J. Petersen, T.J. Menke, B.A. Brody, D.H. Kuykendall, J.C. Hollingsworth, C.M. Ashton, and N.P. Wray. 2002. A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee. *New England Journal of Medicine* 347:81–8.
- Moynihan, R. 2003a. Claims by Charity Exaggerate Dangers of Osteoporosis *BMJ* 327:58.
- Moynihan, R. 2003b. Cochrane at Crossroads over Drug Company Sponsorship. *BMJ* 327:924–6.
- Moynihan, R. 2003c. The Making of a Disease: Female Sexual Dysfunction. *BMJ* 326:45–7.
- Moynihan, R., I. Heath, and D. Henry. 2002. Selling Sickness: The Pharmaceutical Industry and Disease Mongering. *BMJ* 324:886–91.
- Mulrow, C. 1987. The Medical Review Article: State of the Science. *Annals of Internal Medicine* 106:485–8.
- Mulrow, C. 1995. Rationale for Systematic Reviews. In *Systematic Reviews*, edited by I. Chalmers and D. Altman, 1–9. London: BMJ Books.
- Mulrow, C., and D. Cook, eds. 1998. *Systematic Reviews: Synthesis of Best Evidence for Health Care Decisions*. Philadelphia: American College of Physicians.
- Olsen, O., and P. Gøtzsche. 2001. Cochrane Review on Screening for Breast Cancer with Mammography. *Lancet* 358:1340–2.
- Oxman, A., and G. Guyatt. 1993. The Science of Reviewing Research. In *Doing More Good than Harm: The Evaluation of Health Care Interventions*. New York: Annals of the New York Academy of Sciences 703:125–34.
- Payer, L. 1990. *Medicine and Culture: Notions of Health and Sickness in Britain, the US, France and West Germany*. London: Gollancz.
- Payer, L. 1992. *Disease-Mongers: How Doctors, Drug Companies, and Insurers Are Making You Feel Sick*. New York: John Wiley and Sons.

Rossouw, J.E., G.L. Anderson, R.L. Prentice, A.Z. LaCroix, C. Kooperberg, M.L. Stefanick, R.D. Jackson, S.A. Beresford, B.V. Howard, K.C. Johnson, J.M. Kotchen, and J. Ockene. Writing Group for the Women's Health Initiative Investigators. 2002. Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women's Health Initiative Randomized Controlled Trial. *Journal of the American Medical Association* 288:321–33.

Sacks, H., D. Reitman, D. Pagana, and B. Kupelnick. 1996. Meta-Analysis: An Update. *Mount Sinai Journal of Medicine* 63 (May/Sept.):216–24.

Schünemann, H.J., D. Best, G. Vist, and A. Oxman for The GRADE Working Group. 2003. Letters, Numbers, Symbols and Words: How to Communicate Grades of Evidence and Recommendations. *Canadian Medical Association Journal* 169:677–80.

Shea, B., D. Moher, I. Graham, B. Pharm, and P. Tugwell. 2002. A Comparison of the Quality of Cochrane Reviews and Systematic Reviews Published in Paper-Based Journals. *Evaluation and the Health Professions* 25:116–29.

Shulz, K., I. Chalmers, R. Hayes, and D. Altman. 1995. Empirical Evidence of Bias: Dimensions of Methodological Quality Associated with Estimates of Treatment Effects in Controlled Trials. *Journal of the American Medical Association* 273:408–12.

Simes, J. 1986. Publication Bias: The Case for an International Registry of Clinical Trials. *Journal of Clinical Oncology* 4:1529–41.

Upshur, R. 2003. Are All Evidence-based Practices Alike? Problems in the Ranking of Evidence. *Canadian Medical Association Journal* 169:672–3.

U.S. Department of Health and Human Services. 1999. *Mental Health: A Report of the Surgeon General*. Rockville, Md.: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, and National Institute of Mental Health.

Welch, G., and J. Mogielnicki. 2002. Presumed Benefit: Lessons from the American Experience with Marrow Transplantation for Breast Cancer. *BMJ* 324:1088–92.

Wennberg, J., M. Barry, F. Fowler, and A. Mulley. 1993. Outcomes Research, PORTs, and Health Care Reform. In *Doing More Good than Harm: The Evaluation of Health Care Interventions*. New York: Annals of the New York Academy of Sciences 703:52–62.

West, S., V. King, T. Carey, K. Lohr, N. McKoy, S. Sutton, and L. Lux. 2002. *Systems to Rate the Strength of Scientific Evidence*. Evidence Report/Technology Assessment prepared for Agency for Healthcare Research and Quality (Number 47).

Woloshin, S., L.M. Schwartz, J. Tremmel, and G. Welch. 2001. Direct-to-Consumer Advertisements for Prescription Drugs: What Are Americans Being Sold? *Lancet* 358:1141–6.

WEB SITES

<http://hiru.mcmaster.ca/epc/> (accessed Nov. 24, 2003).

<http://www.ahrq.gov/about/profile.htm> (accessed Nov. 24, 2003).

<http://www.ahrq.gov/clinic/epcix.htm> (accessed Nov. 24, 2003).

<http://www.ahrq.gov/clinic/uspstfix.htm> (accessed Nov. 24, 2003).

<http://www.bcbs.com/tec/index.html> (accessed Nov. 24, 2003).

<http://www.campbellcollaboration.org/> (accessed Nov. 26, 2003).

<http://www.ccohta.ca> (accessed Nov. 24, 2003).

http://www.cebm.net/levels_of_evidence.asp (accessed Nov. 20, 2003).

<http://www.clinicalevidence.com/lpBinCE/lpext.dll?f=templates&fn=main-h.htm&2.0>
(accessed Nov. 24, 2003).

<http://www.clinpol.mc.duke.edu/ProjectDir/EvidenceBasedMedicine/EPC/epc.html>
(accessed Nov. 24, 2003).

<http://www.cms.hhs.gov/coverage/> (accessed Nov. 24, 2003).

http://www.cms.hhs.gov/mcd/index_list.asp?list_type=ncd (accessed Nov. 24, 2003).

<http://www.cms.hhs.gov/ncdr/memo.asp?id=7> (accessed Nov. 24, 2003).

<http://www.cms.hhs.gov/ncdr/memo.asp?id=81> (accessed Nov. 26, 2003).

<http://www.cochrane.dk/cochrane/handbook/hbook.htm> (accessed Nov. 24, 2003).

<http://www.cochrane.org> (accessed Nov. 24, 2003).

<http://www.cochrane.org/cochrane/cc-broch.htm#LOGO> (accessed Nov. 26, 2003).

<http://www.cochrane.org/cochrane/revabstr/AB000031.htm> (accessed Nov. 24, 2003).

<http://www.cochrane.org/resources/glossary.htm> (accessed Nov. 20, 2003).

<http://www.cochrane.org.us/centralmn.htm#SC> (accessed Feb. 25, 2004).

http://www.ecri.org/about/AboutECRI_Frameset.asp (accessed Nov. 24, 2003).

<http://www.embase.com> (accessed Feb. 26, 2004).

<http://www.g-i-n.net/index.cfm?fuseaction=links> (accessed Nov. 24, 2003).

<http://www.hopkinsmedicine.org/epc/> (accessed Nov. 24, 2003).

<http://www.inahta.org/> (accessed Nov. 13, 2003).

<http://www.jameslindlibrary.org/> (accessed Nov. 24, 2003).

<http://www.metaworks.com> (accessed Nov. 24, 2003).

<http://www.nice.org.uk/cat.asp?c=137> (accessed Nov. 24, 2003).

<http://www.ohsu.edu/epc> (accessed Nov. 24, 2003).

<http://www.pubmed.gov> (accessed Feb. 26, 2004).

<http://www.rti.org/epc/home.html> (accessed Nov. 24, 2003).

<http://www.sign.ac.uk/> (accessed Nov. 24, 2003).

<http://www.who.dk/eprise/main/WHO/Progs/HEN/Home> (accessed Nov. 13 2003).

<http://www.york.ac.uk/inst/crd/> (accessed Nov. 24, 2003).

THE AUTHOR

Ray Moynihan is a Washington, D.C.-based reporter with an international reputation for quality medical journalism. A long-time radio and television reporter with the Australian Broadcasting Corporation, he has also published original studies, opinion pieces, and investigative reports in the *New England Journal of Medicine*, *Lancet*, *BMJ*, and the *Medical Journal of Australia*. He has also written previously for the Milbank Memorial Fund and is the author of *Too Much Medicine?* (ABC Books, 1998). In 1999 he was a Harkness Fellow in Health Care Policy at Harvard University. His most recent award is for outstanding journalism, given by the Medical Journalists' Association in Britain in 2003. He is currently a visiting editor and contributor to *BMJ*.

SELECTED PUBLICATIONS OF THE
MILBANK MEMORIAL FUND

A complete list of the Fund's reports may be viewed online at www.milbank.org. Single or multiple copies of print editions of each report are available without charge while supplies last. Most reports are also available electronically on the Fund's Web site.

Making Sense of the System: How States Can Use Health Workforce Policies to Increase Access and Improve Quality of Care

by Edward Salsberg

co-published with the Reforming States Group

2003

Available electronically only at

<http://www.milbank.org/reports/2003salsberg/2003salsberg.html>

Food Safety Updated: Developing Tools for a More Science- and Risk-Based Approach

by Michael R. Taylor, Margaret O'K. Glavin, J. Glenn Morris, Jr., and Catherine E. Woteki

co-published with Resources for the Future

2003

56 pages

Available electronically at

<http://www.milbank.org/reports/2003foodsafety/030731foodsafety.html>

Implementing the Resident Assessment Instrument: Case Studies of Policymaking for Long-Term Care in Eight Countries

2003

Available electronically only at

<http://www.milbank.org/reports/interRAI/030222interRAI.html>

September 11 and the Shifting Priorities of Public and Population Health in New York

by David Rosner and Gerald Markowitz

2003

68 pages

Rebalancing Long-Term Care in New Jersey: From Institutional toward Home and Community Care

by Susan C. Reinhard and Charles J. Fahey

2003

Available electronically only at

<http://www.milbank.org/reports/030314newjersey/030314newjersey.html>

2000-2001 State Health Care Expenditure Report

co-published with the National Association of State Budget Officers and the Reforming States Group

2003

92 pages

Proactive Hazard Analysis and Health Care Policy

by John E. McDonough

co-published with ECRI

2002 36 pages

Achieving Better Health Outcomes: The Oregon Benchmark Experience

by Howard M. Leichter and Jeffrey Tryens

2002 60 pages

2001 Robert H. Ebert Memorial Lecture: Health Care Quality and How to Achieve It

by Kenneth I. Shine

2002 24 pages

Informing Judgment: Case Studies of Health Policy and Research in Six Countries

co-published with the Cochrane Collaboration

2001 212 pages

1998-1999 State Health Care Expenditure Report

co-published with the National Association of State Budget Officers and the Reforming States Group

2001 204 pages

When Care Becomes a Burden: Diminishing Access to Adequate Nursing

by Claire M. Fagin

2001 48 pages

Why Health Is Important to U.S. Foreign Policy

by Jordan S. Kassalow

co-published with the Council on Foreign Relations

2001 32 pages

The Origins and Implementation of BadgerCare: Wisconsin's Experience with the State Children's Health Insurance Program (SCHIP)

by Coimbra Sirica

2001 44 pages

Accountability through Information: What the Health Care Industry Can Learn from Securities Regulation

by William M. Sage

2000 76 pages

CALIFORNIA/MILBANK BOOKS ON HEALTH AND THE PUBLIC

The following books are co-published with and distributed by the University of California Press.
For information or to order, call 1-800-822-6657 or visit <http://www.ucpress.edu>.

What Price Better Health? Hazards of the Research Imperative

by Daniel Callahan

2003 341 pages

\$29.95 cloth

When Walking Fails: Mobility Problems of Adults with Chronic Conditions

by Lisa I. Iezzoni

2003 380 pages

\$60.00 cloth; \$19.95 paper

Death Is That Man Taking Names: Intersections of American Medicine, Law, and Culture

by Robert A. Burt

2002 256 pages

\$29.95 cloth

Big Doctoring in America: Profiles in Primary Care

by Fitzhugh Mullan

2002 302 pages

\$29.95 cloth

Deceit and Denial: The Deadly Politics of Industrial Pollution

by Gerald Markowitz and David Rosner

2002 464 pages

\$34.95 cloth

Public Health Law and Ethics: A Reader

edited by Lawrence O. Gostin

2002 536 pages

\$60.00 cloth; \$35.00 paper

Public Health Law: Power, Duty, Restraint

by Lawrence O. Gostin

2000 518 pages

\$60.00 cloth; \$24.95 paper

Experiencing Politics: A Legislator's Stories of Government and Health Care

by John E. McDonough

2000 336 pages

\$50.00 cloth; \$19.95 paper

The Corporate Practice of Medicine: Competition and Innovation in Health Care

by James C. Robinson

1999 306 pages

\$45.00 cloth; \$19.95 paper

The Fund also publishes the *Milbank Quarterly*, a multidisciplinary journal of population health and health policy. Information about subscribing to the *Quarterly* is available by calling toll-free 1-800-835-6770, or at www.milbank.org/quarterly/.

Information about other work of the Fund is available from the Fund at 645 Madison Ave., 15th Floor, New York, NY 10022, (212) 355-8400. Fax: (212) 355-8599. E-mail: mmf@milbank.org. On the Web: www.milbank.org.

Copyeditor:

Marie Shear

Design and Typography:

The Boland Design Company

Printing:

Prestone Printing Company, Inc.